

European setback at Milan

Disagreement among members of the European Communities last weekend at Milan will hamper the cause of technical collaboration. Modest plans are now in order.

ALL the high hopes of the past several weeks that Europe would find a way of making itself technologically competitive with the rest of the world have been thrown into limbo. That is the simplest reading of the chaos into which the affairs of the European Communities have been plunged by last weekend's summit meeting at Milan. Formally, the heads of government did manage to agree that something should be done to put flesh on the bones of the French Eureka project; the French government is to call a meeting of foreign and research ministers in the second half of this month at which they will attempt to decide how the Eureka enterprise should work. The present plan is that governments other than the twelve members and putative members of the Communities should be free to participate if they wish. (Austria, Sweden and Switzerland are likely to be welcomed as outsiders.) The meeting is not intended to choose projects for collaboration but, rather, to define a mechanism which, governments agreed at Milan, should be above all flexible. The obvious snag is that while European governments are at loggerheads among themselves about the political future of their community, agreements on even technological collaboration are bound to seem unreal.

Ironically, the fiasco at Milan appears to have sprung from a sensible if belated recognition by European governments that the time has come to make what is often called the common market operate as such. Europe cannot be surprised that it is technologically often at a disadvantage in comparison with Japan and the United States when the present restraints on trade and competition persist. During the past few months, the European Commission has been hard at work on the definition of the steps that must be taken to dismantle these artificial and often illegal impediments to efficiency. Mercifully, all governments appear to agree that getting rid of these restrictions will be impossible while member states hang onto the principle that all decisions about the working of the Communities must be reached unanimously (or that any dissident member may veto an unpalatable proposal, as West Germany vetoed last month's agreement that cereal prices should be reduced). This is why the British went to Milan pushing for a package of procedural reforms that would have restricted a government's right to use the veto, and otherwise have made the community's decision-making more orderly. France and West Germany, on the other hand, with the unexpected backing of Italy, pushed instead for a revision of the Treaty of Rome, the legal instrument which is the foundation of the European Economic Community. The outcome is a fuzzy decision that, later in the year, there should be a formal conference to consider amendments to the treaty. This is a recipe both for delay and for further quarrelling within the community, if only because governments represented at this meeting will be able to veto any of the decisions it may reach. Especially because there is as yet no word on when or how the

conference will be organized, the upshot is bound to be that no framework for technological collaboration in Europe can be assured by the time of the Eureka meeting later this month.

Worse than that, the Milan meeting cannot but further sour relationships between European governments. Only occasionally in the past decade does it seem to have been possible for the heads of the European governments to meet without uncovering some obdurate disagreement. Last weekend's fracas at Milan will have further emphasized the degrees to which different members of the European Communities harbour different but serious reservations about the enterprise to which they belong. The European Communities have survived for the past thirty years chiefly because there has been a general recognition that attempts to force the pace towards European integration will be self-defeating if individual members are faced with politically unpalatable decisions. The attempt that will be made, later in the year, to force through amendments to the Treaty of Rome is bound to throw sand into the machinery of all kinds of European collaboration. Eureka included.

What can be salvaged from this mess? The best way forward for Eureka will be to postpone final decisions about the content of the programme of research and development on which Europe collectively seems eager to embark. In due course, it may emerge that there are technical projects of such originality that they promise opportunities for particular European industries to become genuinely competitive. Schemes for attempting to beat IBM at the game which it has played successfully over past decades should not, at this stage, qualify. It should also be clear that Eureka should not become one of those ambitious community projects in which there is some central pot of gold contributed by governments for a share of which companies in member states are encouraged to compete. The best strategy will be to use such public funds as there may be for supporting to provide the infrastructure from which novel technology may spring. Although President Mitterrand's clear wish (and political need) to clothe Eureka ambitiously would not be satisfied by, for example, public support for research and teaching institutions that could make good deficiencies in Europe, that may be the wisest course after the disaster at Milan. □

Where the onus lies

British science should make the government shoulder responsibility for political decisions.

ON the working hypothesis that the British government will not increase the funds earmarked for research in the next financial year (beginning in April 1986), researchers had better take what steps they can to save their own skins. In the next few months, the burden will fall chiefly on the Advisory Board for the Research Councils (ABRC), which technically exists to advise the Department of Education and Science on how the inadequate funds available should be divided. Two immediate issues must be settled. ABRC has to say how it would have the government respond to the report of the Kendrew Committee, which advocated a fortnight ago that Britain should pull out of European collaboration in high-energy physics (chiefly through CERN) unless there can be negotiated a 25 per cent reduction of the

Biological manuscripts

Contributors are reminded that, with the transfer of the Biological Sciences Editor to the Washington office, it will be helpful if they will in future send **four copies of all manuscripts** offered for publication, either (as at present) to London or Washington.



common budget after 1991. Then, unexpectedly, a committee of the Royal Society on public support for geophysics (see *Nature*, 27 June, p.709) has unexpectedly thrown up a series of questions about the Natural Environment Research Council (NERC) which cannot be ignored until the council itself decides the time has come to answer them. What should ABRC say?

The Kendrew strategy is unhelpful for the simple reason that it will yield no budgetary relief when that is needed most, which is now. Even if savings of the order of 25 per cent are possible, which is improbable when CERN members such as Italy are eager to see high-energy physics in Europe expand, the first small relief for the British budget would come only in 1988. The difficulty into which the whole Kendrew exercise was boxed is that Britain is committed to the present scale of spending at CERN for at least the next three years by the undertakings given by the Science and Engineering Research Council (SERC) that there would be no backsliding while the new accelerator is being built, by similar undertakings to the international collaborations building detectors for the new instrument and by the implicit, but equally important, undertakings to graduate students whose careers are predicated on the completion of projects such as these. If Britain now follows Kendrew's advice, the course of events during the next two years is easily predicted. There will be a long wrangle at Geneva about future budgets, Britain will be blamed for rocking what has hitherto been a united enterprise, but in the end (if the strategy is kept to) will be forced to withdraw in humiliation. The political damage done will be no less than if it were decided to pull out of CERN immediately, moral obligations notwithstanding.

This dilemma is not of the research community's making, nor is it one that the community should on its own resolve. The difficulty has arisen because, while the British government refuses to recompense the research councils for the fall in the value of sterling except on an *ad hoc* basis, the CERN subscription (now £35 million a year) has become conspicuous. (No doubt it is only a matter of time before some other committee is appointed to enquire into the cost of the British contribution to the La Palma observatory, inaugurated only last weekend.) ABRC should now firmly throw this problem at the government, which did at the outset of the Kendrew inquiry say that it would have the final say on any proposal to withdraw from CERN. Now is the time for that. ABRC should warn the government of the political risks of pulling out, should explain that it cannot ask members of the scientific community to embark on a probably futile negotiation when the execution of the implied duress is beyond its own control, and meanwhile should insist that this and the other politically sensitive international subscriptions which are at present part of the science budget (such as the SERC subscription to the European Space Agency) are taken off the books of the research councils and handled centrally, by ABRC itself, with proper safeguards against currency fluctuations. Failing a satisfactory reply, ABRC should decline to follow the Kendrew strategy, but instead ask SERC to fight for economies at CERN, to cut back on what it spends domestically on high-energy physics but not to raise the threat of withdrawal until more is known of the future pattern of spending. More constructively, the question might be rather whether the CERN agreement should be amended. Is it strictly necessary that national subscriptions should be determined only by gross national product?

The problem of NERC is also urgent. The argument that this research council, cobbled together in the 1960s, should now be broken up is strong (see *Nature* 311, 493; 1984) but is probably unpalatable to too many influential people. Last week's Royal Society document was only incidentally an inquiry into the research council's way of working, and as such is not sufficient basis for root-and-branch reform. But a few things should be clear. First, it is absurd that Britain should pull out of the Deep Sea Drilling Program because NERC's funds are otherwise committed. ABRC should issue the appropriate instruction. Second, it is a tragic waste of people that the British Geological Survey's staff should this year lack the funds to do scientific work

for similar reasons. The cause may be the unwillingness of other government departments to pay for commissioned research, but the Royal Society's committee is right to say that the British national interest requires continuity in this field. If the survey's work can be secured only by transferring it to some other branch of the British government, so be it. Finally, given the committee's cogent criticisms of NERC's record on the award of research grants to academic geophysicists, there is a strong case for transferring that part of its function lock, stock and barrel to SERC. This is what ABRC should be saying, and quickly. □

Waxman's way

The US Congress should abandon its scheme to set up two new research institutes.

THE US congressional calendar is once again burdened by Representative Henry Waxman's attempt to wish on the National Institutes of Health (NIH) two entirely unnecessary research institutes, one for arthritis and one for nursing. The outcome is predictable, and will be no different from what it has been in the past three years. Waxman's bill, which has already been passed by the House of Representatives, will in due course be sent off to the White House and vetoed by the President. This is what happened last year, and nobody expects the outcome to be any different this time. The explanation is simple enough. The administration thinks the two extra institutes unnecessary, and has said so. It is also suspicious of Waxman's motives. Although this year's version of the bill (like last year's) is a considerable retreat from the earlier versions that would have given Congress detailed control of the way the individual institutes of NIH spend their budgets, there is no reason to think that Waxman has abandoned his ambitions to strengthen congressional control of the NIH budget. But while this legislation lies in limbo, three of the institutes (including the National Cancer Institute) will have to limp along from year to year by means of a continuing resolution that gives the administration authority for a year to keep on spending at last year's rate.

This is hardly the way in which the US Congress, which over the years has been generous almost to a fault to NIH, should deal with its favoured dependants. But it is not likely that, if NIH is required to set up further institutes, the flow of funds from Congress will be increased? That is how supporters of Waxman-like legislations have argued in the past, urging that NIH should welcome the creation of one institute for each recognizably distinct disease. It goes without saying that the political benefits for those members of Congress who force on NIH institutes dedicated to the cure of particular diseases will be far from unimportant. But this suggestion that a symbiotic relationship between Congress and NIH would ultimately benefit both is offensive because it betokens a thorough misunderstanding of the way in which NIH functions.

Congress, its constituency in the United States and biomedical research internationally have hugely benefited from the flexibility with which NIH are organized. In the recent past, the contribution of people at the National Cancer Institute to the understanding of Acquired Immune Deficiency Syndrome (AIDS) would have been no greater if Congress had decreed four years ago that there should be a special institute devoted to this novel scourge. Nor would matters have been much improved if, using the powers to control institute budgets in detail which featured in earlier versions of Waxman's bill, Congress had insisted that teams at the National Cancer Institute should not let their curiosity wander beyond the field of potentially oncogenic viruses so as to encompass that which now appears responsible for AIDS. In short, the recipe of one institute for one disease is a recipe for rigidity in a system where flexibility has miraculously flourished, sometimes against the odds. That is what Mr Waxman, and his colleagues in the Congress, should now openly recognize. □

Soviet scientist

Nuclear winter expert vanishes without trace

Stockholm

A SOVIET computer scientist, who has played a leading part in the nuclear winter debate, disappeared from a meeting in Toledo, Spain, last March, and is now feared by Western colleagues to be dead. Vladimir Aleksandrov was a Soviet delegate to SCOPE (the Scientific Committee on Problems of the Environment) and was an active participant in the SCOPE working party on the environmental effects of nuclear warfare.

Reports of last sightings of Aleksandrov vary, but according to one account he expressed his intention of going back to the meeting centre and was never seen again. According to another source, he was last seen being bundled into a car by several men. His personal effects were later found to have gone from his room, but his passport was found in a litter bin.

Former colleagues at the Lawrence Livermore Laboratories in California report frantic messages from Mrs Aleksandrov, who was greatly surprised and concerned at her husband's failure to return.

Aleksandrov was the author of the Soviet computer model using a two-layer model of atmospheric circulation which was presented in Stockholm in November 1983, and which supported the conclusions of Sagan and Ehrlich on the climatic aftermath of a nuclear war. Aleksandrov had allegedly told a Western colleague that he had been told to "get busy" on nuclear winter two months before the Sagan "peer review" meeting at the American Academy of Arts and Sciences in Boston in April 1983.

Aleksandrov had, during the past eighteen months, gradually been replaced as the chief Soviet scientific spokesperson on nuclear winter by Kiril Kondrat'ev, whose papers concentrated on nitrous oxides and trace gas effects, and paid little heed to smoke and dust. This may mean nothing more, however, than that once the computer model was complete, the Hydrometeorological Institute of the Soviet Academy took over the main burden of the nuclear winter campaign.

The silence of Aleksandrov's Western colleagues is difficult to explain. Some, however, seem to have thought that he had defected, and that the car incident was a cover-up. These have expected him to emerge from a safe house in some Western country in due course. But in that case, protest from the Soviet Union about what could well be interpreted as a kidnapping would have been expected, but none has been forthcoming. Nor have there been attempts to brand Aleksandrov

as a defector. One source which claims to be in contact with several Western intelligence services says that it is "not on the cards that the West has him".

If Aleksandrov was repatriated to the Soviet Union by force, as some of his friends now fear, and he was killed either deliberately or in the course of such an attempt, it is not clear what the motive

could have been. One suggestion is that he was feared to be a defection risk. Another is that Aleksandrov had become convinced that the Soviet concern with nuclear winter was merely for public and international opinion and that the Soviet military planners were working on the basis of a "survivable level" of fall-out, and that he intended to voice these views.

Such suggestions, perhaps farfetched, reveal the growing concern about his fate. Hopefully, however, even at this late date, the Soviet Union or some Western government will respond with some explanation and, if possible, the production of Aleksandrov in person.

Vera Rich

Nuclear power

US plans to set industry free

Washington

ATTEMPTS are again being made to revitalize the moribund US nuclear power industry by speeding up the licensing of new plants. Three bills* now before Congress, from the Department of Energy, the Nuclear Regulatory Commission (NRC) and from the industry itself, seek to encourage the use of standard plant designs by giving NRC authority to issue combined construction and operating licences (COLs) at the design stage. At present, NRC must hold a second examination of safety issues before a completed plant can be operated.

The argument is that by forcing all safety issues into a single examination, NRC will give utilities an incentive to perfect complete designs that can be adapted to the requirements of different sites. Construction permits have often been issued for incomplete plans, an approach characterized by the Union of Concerned Scientists (UCS) as "design-as-you-go", resulting in frequent expensive modifications.

There is widespread agreement that the present two-step licensing procedure is inefficient, and that engineering expertise would be better used if there were but a few standard plant designs to be examined. But opponents of the new proposals argue that they fail to encourage standard designs, and that they make it harder to raise serious safety problems once a plant has been built.

Each of the bills recognizes that some kind of pre-operational safety review is necessary before operations can begin, but to prevent determined opponents of a plant from raising questions of safety that have been adequately considered at the design stage, each contains some restriction on the issues that can be raised in a pre-operational review. The NRC bill, for example, says a new issue must be a substantial dispute of fact that cannot be resolved except at a hearing, and that it must either concern non-compliance with a previously-issued COL or a question that

could not previously have been examined. Controversially, the NRC bill also says that for a post-construction modification to be required, it must substantially improve the overall safety of the facility. Overall safety is examined using probabilistic risk assessment, which according to its critics is nothing better than an obfuscatory device that, by being easily manipulable, is used to give spurious authority to safety claims. The NRC bill also raises hackles because it would allow a plant to begin operating before hearings end.

The Department of Energy's bill would make it even harder for safety issues to be introduced retrospectively, as it would allow NRC to make conclusive design approvals that would remain in force for 10 years no matter what information was subsequently discovered; furthermore, it would allow the licensee to deviate from the design in the COL as the licensee saw fit. The industry bill gives intervenors greater latitude to introduce safety issues at a late stage, but requires NRC to perform a central review before incorporating changed requirements. This could mean delays of years.

Opposition to the three bills, voiced more articulately by UCS, focuses on the argument that NRC already has the legal authority to approve standardized plant designs. UCS says it is the fault of the industry that such designs have not been used in the past, because it has sought construction permits on the basis of an examination of very incomplete plans. According to UCS, the depressed state of the industry — no new plants ordered since 1978 — would change if NRC insisted on seeing a complete design before granting construction approval. Industry spokesmen contritely acknowledge that approval of incomplete plans is counter-productive and promise that utilities will submit "essentially complete" plans in future.

Tim Beardsley

*HR 1029 (from the industry) and HR 1447 (from NRC); the Department of Energy bill has not yet formally been introduced in the 99th Congress.

Biotechnology

Variety at *Nature* conference

PLANTS engineered to produce antibodies against viral pathogens, protein frameworks into which various of functional units can be plugged at will and biosensors based on a marriage of immunology and fibre optics were among the more engaging ideas to emerge from *Nature's* "New Technology in Biotechnology" conference at London's Novotel last week.

Clearly the path to commercial reality from an engaging biotechnological idea is long and exceedingly stony, but it is increasingly travelled. Ten years after the discovery of hybridoma technology, John Birch could talk of the production at Celltech of around a kilogram of monoclonal antibody a year, mostly for blood group testing, (see *Nature* 27 June, p.705) and its sale for, typically, £1,000 a gram.

Clever variations on the basic theme of monoclonal antibodies are already proving their worth at the Molecular Research Council's Laboratory of Medical Biology in Cambridge. Most are the result of engineering the genes for antibodies, replacing an inessential part of the sequence with another gene. The results are hybrid, bifunctional proteins, several of which were described by Michael Neuberger. In one, the substitute gene is for an enzyme used in DNA sequencing. After the hybrid antibody/enzyme is produced in large quantities by standard hybridoma technology, it is very simply and efficiently immunopurified by means of its antibody end; the other end is now supplying the laboratory's considerable demand for the enzyme. On the other side of the coin, a protein that is unable to induce good monoclonal antibodies has been engineered to contain an extra short sequence so that it can be recognized by an existing and excellent monoclonal antibody against the short sequence.

Daniel Thomas of the University of Compiègne foresaw the value of hybrid enzymes in second-generation immobilized enzyme technology. For example, the lifetime of an oxidase that is susceptible to damage by some of the oxygen radicals it produces could be lengthened if the ox-

idase were to be hybridized with a superoxide dismutase, which can consume the radicals. Other important second generation systems would, for example, have built-in systems of co-factor regeneration on electrochemical surfaces or would operate in organic solvents, perhaps to reverse an enzyme reaction or in organic synthesis. More sophisticated protein engineering will involve the subtle redesign of active sites of enzymes, which will remain very much a matter of trial and error until much more is known of the determinants of protein folding, according to Greg Winter of the Medical Research Council Laboratory of Molecular Biology in Cambridge.



Even without engineering, enzymes and antibodies have many projected uses in biosensors, very few of which have yet progressed from the laboratory to the market place. Michael Flanagan of University College London described prototype optical immunosensors in which the binding of fluorophore-labelled antigen to antibodies immobilized on a surface of the device is measured by means of evanescent light waves of plasmon resonance laser light. For potentiometric devices, enzyme immobilization techniques badly need adapting to suit the surface of microelectronic devices, he said. For amperometric devices, John Higgins of Cranfield Institute of Technology and Leicester Biocentre described current interest both in the organometallic ferrocenes as mediators of electron transfer to electrodes from enzymes bound to them and in a newly discovered class of dehydrogenase enzymes that use pyrrolo-quinolinequinone as a cofactor in place of the more usual but problematic nicotinamide adenine dinucleotide.

Recombinant DNA technology lags behind that of hybridomas and perhaps even biosensors in its commercialization. But there are products in the offing. Pierre Tiollais of the Pasteur Institute described the advanced state of a hepatitis B vaccine from engineered mammalian cells and Art



Levinson of Genentech claimed that the company's tissue plasminogen activator has been able to dissolve blood clots and restore blood flow through the coronary arteries of some patients in a clinical trial.

For plant breeders, the most immediate gain is likely to be strains that have been engineered to be herbicide resistant, according to Jozeph Schell of Ghent University and the Max Planck Institute for Plant Breeding Research in Köln. Less advanced are attempts to improve potato protein quality and quantity, to transfer light-inducibility to new genes and to engineer cereal crops. And but a glint in Schell's eye is a plan to give tobacco plants the mammalian gene necessary to produce a protective antibody against tobacco mosaic virus. For animal breeders, Ralph Brinster of the University of Pennsylvania offered encouraging news of the general applicability of the introduction of genes into fertilized eggs and the tissue-specific expression of the genes at high levels in the subsequent transgenic animals. One recently learnt fact is that the removal of residual bacterial sequences from a gene construction introduced into the eggs can greatly increase the gene's expression in the tissues of the transgenic animals.

Peter Newmark

● For those who would like a record of the *Nature* conference, cassette tapes are available from: Audio Bass Ltd, 16 Longland Drive, Totteridge, London N20 8HE, UK (Tel: 01-445 6776 or 01-349 4041). Please quote programme number 064-85.

● The correct answers in the exhibition competition were as follows:

1. Tissue culture bottle/Laminar flow hood
2. Eppendorf microfuge
3. Dilon pipetman
4. Scintillation counter
5. Tube sealer
6. Fibre optics

None of the entrants identified all the photographs correctly. Sally Roberts and Pam Jones, both of the Laboratory of Molecular Biophysics at Oxford, got five out of the six right, and in the draw Sally Roberts was selected as the winner of the personal computer. □



French education

Higher marks for effort

ONLY 37 per cent of French children attain the coveted *baccalauréat* (*le bac* for short) which guarantees entry to university. But the figure should be 80 per cent by the end of the century, according to the French minister of national education, M. Jean-Pierre Chevènement. Such an increase would raise the number of students in higher education in France to 2 million, four times the figure in Britain, which has a population of the same size.

If all 2 million students were to proceed to higher education, the 80 or so French universities could not cope. At present, with just one million students, as many as two-thirds of entrants leave university after the first year, without a diploma. (Lecturers dub them "ghost students" — some do not even attend courses.) In practice, the ministry of education expects that only a fifth of the extra million French students, who will come through a new form of baccalaureate, *le bac professionnel*, will be seeking places in universities or the French technical colleges, the *Instituts Universitaires de Technologie* (IUT), which are sought after by parents and students because education there usually results in a job.

Others may go to new universities of technology (there is only one at present in France, at Compiègne) or to an "open university", both of which have been promised in recent weeks by the French government. But most of those attaining the new *bac professionnel* are expected to go straight into employment.

The new scheme is designed to loosen up the less academic and employment-oriented French high schools, to make them adapt their education to the real needs and demands of local employers, and to extend the extra 2–3 years of secondary education required for *le bac* (the examination is usually taken at the age of 18) to a wider social group.

The new move must also be seen in its true context as part of the phenomenon of the minister himself. Chevènement is an old-fashioned socialist working in the 1980s. He believes in equality of opportunity: the new *bac* will provide that. He believes in industry: the new *bac* will be tuned to it. But secondary school students will have to work. Chevènement believes in attainment. He also believes in science, as he believes passionately that socialism is equated with knowledge. As President François Mitterrand's first research minister, Chevènement reversed the anti-industry attitudes of most of the French research community, which may be the most substantial achievement of the socialist administration in its four years in power. And now, after a period in political limbo, he seeks to make further lasting marks on France by reshaping the whole education system.

Out have gone the "advanced" liberal attitudes of the 1968 student revolution, whose anti-authoritarian spirit affected schools almost as much as universities, and in has come what the minister calls "republican elitism", basically a renewed respect for an encouragement of academic attainment and for the republican ideals of post-revolutionary France.

"A structured school is more democratic than a lax school", Chevènement wrote recently, discussing the nature of his "elitism". In a lax school, the more confident and middle class pupils rise to the top of the heap. In a structured school, native ability wins through. The proportion of

working class children at universities is still very low, and the minister intends to change that by providing academic and professional opportunities. Hence the new *bac*, which will also have the advantage of helping to train the kind of technicians that new industries need.

Among other important innovations, Chevènement is working on reducing the power of the mathematicians in French schools, which is thought to discourage experimental and observational sciences, in particular biology and geology.

Chevènement has already outlined his ambitions for the universities (see *Nature* 9 May, p.88) but he may not have long to achieve his aims. There will be a general election next year, and the polls suggest that the socialists will be defeated.

Robert Walgate

European computing

Toulouse carries ambitious torch

THE FRENCH proposal for a European computing centre at Toulouse in south-west France, described by a British computer scientist as "a kind of CERN [the European Organisation for Nuclear Research] for computing", has met at least with the approval of Italy. But Britain appears sceptical, while in Toulouse itself, hopes were high for support from West Germany.

That was the position last week over the future of CERFACS, the "Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique", an £18-million-a-year institute for research and training in numerical analysis whose establishment has been championed for over two years now by a group of Toulouse astrophysicists, aerospace engineers and politicians. According to CERFACS supporters, existing techniques of numerical analysis are out-of-date, having been designed for machines with a single central processing unit. With new computer architectures involving vectorial and, increasingly, parallel processing, where the same or different functions can be performed many times simultaneously, numerical analysis needs to be rethought. CERFACS would do the rethinking, provide training and help apply the techniques to develop areas such as the creation of numerical "wind tunnels" (flight testing by pure hydrodynamical computation).

According to CERFACS supporters, with an eye on the 1986 French budget as it approaches its final month of inter-ministerial bargaining, apart from possible support from Germany and Italy, the Netherlands, Denmark and Switzerland are also ready to join in, provided a second country (other than France) offers support. And in France, while government support is still not guaranteed, there is a chance that CERFACS may be caught up in the political snowball of "Eureka"

(see pages 8 and 9).

But there is some concern in Paris, and even more in London, that CERFACS is too ambitious and costly, and should be prepared to grow "at a natural pace". There are "a lot of traps" in the Toulouse plans for CERFACS to start up in September 1987 with over 100 staff, rising to nearly 200 by 1991, it is said in Paris. In London there is a feeling that the centre would not offer "value for money".

Professor Jean Yoccoz, the director of the French nuclear and particle physics institute IN₂P₃, however, says "there are always pluses and minuses" and that "the best of CERFACS" is its plan to mix industrial and fundamental research on one site, as well as training. And Toulouse, with its aerospace and robotics industry, "is a good place for that, at least for France".

A leading British computer scientist, however, is less impressed. He feels the emphasis on new numerical and software techniques for use with vectorial supercomputers is "very esoteric" and "not leading edge computing science".

In Toulouse, however, these criticisms have been met with "surprise". CERFACS is prepared to deal with whatever computer technology is in the forefront, be it a vectorial Cray 2, or the German multiprocessor Suprenum, or a machine still on the drawing board. Far from being esoteric, CERFACS — or something like it — would be essential to the proper use in Europe of these new machines, it is claimed in Toulouse and whatever equipment was provided would moreover be networked to users throughout Europe. "We aim to respond to user needs", a spokesman says. And although the planned rate of development of CERFACS is indeed ambitious, Europe cannot afford to wait to be pre-empted in yet another area by the United States and Japan.

Robert Walgate

Deep-sea drilling

Britain squeezed onto margins

Washington

STRAINS are beginning to show over Britain's continued inability to find the \$2.5 million membership fee needed to join the Ocean Drilling Program (ODP). British scientists have been removed from all planning and advisory committees of the programme, and according to Robert Kidd, manager of scientific operations of the research vessel *Joides Resolution*, British scientists are unlikely to be invited to participate in future cruises.

Britain at present enjoys observer status in ODP, which permits it only to send an observer to meetings of the executive committee of Joint Oceanographic Institutions Deep Earth Sampling (JOIDES). Japan, which also has observer status, recently signed a memorandum of understanding with the United States saying that it will join the programme as a full member at the start of October this year; no such guarantee has yet been made by Britain. As a result, the only British scientist on the current cruise of *Joides Resolution* in the Norwegian Sea, Lindsay Parsons of the Institute of Oceanographic Sciences, is likely to be the last until Britain rejoins, unless there is a cruise in British territorial waters.

Planners at ODP find themselves in a political bind: although they want to invite individual British scientists because of their expertise in particular fields, it is felt that it would be unwise to send the British government a signal that its best scientists will be able to participate in ODP whether or not Britain becomes a member. The assistant programme director of ODP at the US National Science Foundation, Alexander Sutherland, says JOIDES has been "pretty adamant" that British scientists should not be offered guaranteed places, out of fairness to participating countries. Dr Parsons' participation in the current cruise, leg 104, was arranged before the JOIDES committee forbade guaranteed places for British scientists, and was supported (despite misgivings in some quarters) by the two co-chief scientists on the cruise, Olav Eldholm of the University of Oslo and Jorn Thiede of the University of Kiel.

Britain's difficulties in finding the necessary money has particularly annoyed Canada, which has paid despite having a much smaller gross national product. Although the European Science Foundation and Australia may join ODP as a consortium, JOIDES has ruled out the possibility of Britain joining as part of a consortium.

Britain's Natural Environment Research Council still hopes to be able to raise the necessary funds in time, although many people in the United States seem resigned to the fact that Britain will not be a member next year, at least. One political

factor that may yet carry the day in Britain, however, is a planned cruise in the Weddell Sea: a scientific interest and

presence in Antarctica would undoubtedly bolster British negotiating positions when the Antarctic Treaty comes up for renegotiation at the end of the decade. Only by being in at the planning stages will Britain be able to influence the shape of the cruise.

Tim Beardsley

Auto emissions

Europe agrees at last

Brussels

AFTER an all-night session, environment ministers of the European Economic Community (EEC) came to the close of their meeting in Luxembourg on 28 June just in time for breakfast with agreement on new tougher European exhaust emission controls, the first of which will be applicable from 1988.

The deadlock over this controversial issue, which has kept the environment ministers busy over one aspect or another for two years, was overcome by a decision not to follow the US example, despite pressure from West Germany, to quantify the exhaust emission limit for nitrogen oxides (NOX) on their own for cars in the intermediate range of 1.4 to 2.0 litres.

The combined figure for nitrogen oxides and hydrocarbons (HC) has been common practice until now in the regular modifications made to the 1970 EEC directive on exhaust emissions, and gives motor manufacturers room for manoeuvre within a range of engineering options. In getting to grips with emissions, one of the problems in engine design is that modifications to the combustion process aimed at reducing, for example, hydrocarbon emissions increase the production of NOX and *vice versa*.

The limit finally agreed — 8 gram/test for HC and NOX together, the European Commission's proposal — falls half-way between what West Germany wanted on the one hand and what the British were prepared originally to accept on the other.

For the United Kingdom, the difficulty lay in accepting a figure that it might be impossible to obtain using lean-burn single-point ignition, under development in the United Kingdom. Furthermore, the use of this technology (lean-burn plus oxidation catalyst) had been part of the agreement hammered out by environment ministers at their previous meeting in March. As it turned out, Mr William Waldegrave, Under Secretary of State at the UK Department of the Environment, agreed on his own responsibility to the

figure pending a government decision in London. No hitches are expected.

This spirit of European compromise has meant less of a warm welcome from environmentalists in West Germany for their interior minister Friedrich Zimmermann, for whom coherence in the Community's external market, at the end of the day, outweighed tough political pressure on a national level.

Ministers also accepted a number of other ancillary points, said to have been initially proposed by French environment minister Huguette Bouchardeau, which aim at further reducing exhaust pollution (and tension between the member states):

- That small cars (under 1.4 litre) would have to conform earlier to stricter standards, bringing them in line with US environmental equivalent effects, by 1992 for new models and 1993 for all new cars.

- That the United Kingdom and France would agree to drop complaints made to the European Commission against the national fiscal incentives for "clean cars" announced by West Germany at a time when European legislation on the subject was being prepared (illegal under Community law).

- That the Commission should propose legislation in the next three months to limit emissions from lorries, and in particular on diesel particulates (see *Nature*, 311, 401; 1984). Ministers also laid down new standards for diesel engines with a capacity of 2 litres and over. While, in the past, limits for diesel engines have had the advantage of a 25 per cent waiver in their favour, they will in the future have to comply with the same new limits established for the intermediate range of conventional car engines.

- That the Commission should make proposals on speed limits and on a regular monitoring system on vehicle compliance, possibly based on the tough system applied in Belgium.

In addition, the Commission itself may decide to propose further measures.

Anna Lubinska

	Emission standards		Deadline (1 October)	
	CO	HC + NOX	New models	New cars
Cars over 2 litres	25	6.5 (3.5 NOX)	1988	1989
Cars of 1.4-2 litres	30	8.0	1991	1993
Cars below 1.4 litres	45	15.0 (6.0 NOX)	1990	1991

(Stage 1)

(Stage 2 to be agreed in 1987 with implementation by 1992-93.)

Japanese education

Council's criticisms start row

Tokyo

JAPAN's special Council on Education, set up to prepare a blueprint for a massive reform of the education system, delivered its first set of recommendations last week to a chorus of complaints and disapproval.

It is only ten months since the council, with its 25 members all personally picked by Prime Minister Yasuhiro Nakasone, began its activities, so it is perhaps not surprising that it has been unable to reconcile all the fiercely held views on this emotive subject. But the report is more of a philosophical declaration than a practical recipe for reform.

The report castigates the present education system in terms which, if they had been written by a foreigner, would probably have given rise to complaints of anti-Japanese prejudice. The system is accused of producing "too many stereotyped persons without strong individuality" who are not "able to think and judge on their own" because they have been taught only to "memorize information and facts". "Bullying in school, pupils' absenteeism, violence at school and juvenile delinquency" are also described as "grave" and as products of the present system, although Japanese schools are havens of peace compared with those in the West.

To remedy this situation, strong emphasis is placed on the fostering of "individuality". But although this Western notion is repeatedly endorsed, far more column inches of the report are given over to the need to reinforce distinctively Japanese values: children "must love Japan", maintain "a deep understanding and love of Japanese culture" and have "identity as Japanese". These values were, of course, inculcated in the education system that existed before the Second World War, which, indeed, comes in for some oblique praise, while post-war education is criticized for putting "a disproportionate emphasis on rights instead of duties". To the Socialist and Communist parties (the second and fourth largest parties in the Diet) such statements are like a red rag to a bull and they have attacked the report furiously, pointing out the difficulties in a system which is meant to teach people to "think independently" while "fostering Japanese who will inherit the traditional culture".

Philosophical contradictions aside, the report is firm in identifying the real villain behind educational woes: a society that gives too much preference to graduates from leading universities. Because top corporations and ministries recruit from just a few universities, the whole educational system has become oriented towards success in the university entrance examination "hell". In that examination, all that counts is overall score in multiple choice questions; the higher one's score

the better the university (or faculty) one can enter, regardless of vocation and aptitude. The report makes an appeal for a change in attitudes away from a society oriented towards "academic record", but clearly such changes are beyond its power to effect.

A more direct attack is made on the university entrance examination. A new first-stage unitary examination is proposed for use by all universities, public and private, to replace that now used only by public universities. The burden on students, who usually try more than one kind of examination, would thus be eased.

To try to ease competition further down the system, new kinds of schools are proposed that will fuse junior and senior

high-school. Entrance examinations for high-schools have recently become almost as fraught as the university entrance examinations themselves. It is suggested that some such schools work on a credit system for high-school graduation rather than rely on intensive examinations to increase the flexibility of the system.

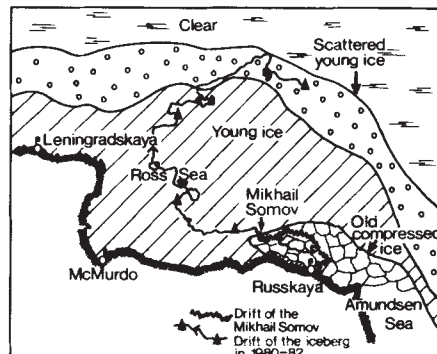
The Council on Education will now soldier on for a further two years, trying to build on this somewhat shaky start. The Prime Minister is giving full backing to the council, and a cabinet-level group is expected to be formed to put its weight behind its recommendations. So far, however, teachers are opposed, the Japan Teachers Union having sworn to fight the reforms "to the end" a few days before the report was even published. The teachers' main concern is to get average class sizes down to 35 from present levels which are closer to 40 pupils.

Alun Anderson

Soviet Antarctica

Trapped vessel in good shape

THE Soviet research vessel *Mikhail Somov*, which has been trapped in the Antarctic pack-ice for more than three months, could eventually drift to safety along a route mapped by a previous expedition aboard the same vessel. Accord-



ing to Candidate of Sciences Al'bert A. Romanov, of the Arctic and Antarctic Research in Leningrad, the ship's present course virtually coincides with that of an iceberg, mapped by radio beacon as part of a Franco-Soviet experiment in 1980-82. The tagged berg drifted on a semi-circular course for more than 3,400 nautical miles before emerging into the warm waters of the Pacific.

The planners of the 30th Soviet Antarctic expedition, of which *Mikhail Somov* is the flagship, do not, however, intend to leave the crew and research personnel trapped for so long. Some 77 people have already been flown out by helicopter, leaving 53 on board for essential ship maintenance and observations of conditions in this previously unexplored sector of the Antarctic. (Conditions are severe, with winds of 17 metres per second, temperatures of -25°C, poor visibility, snowstorms and fog.) Meanwhile, a special taskforce has been established under the deputy head of the state committee for

hydrometeorology and environmental control to "ensure a rapid solution" of the "*Somov* problem". The icebreaker *Vladivostok* of the Far Eastern Shipping Line has been sent south, and is expected to approach *Mikhail Somov* at the beginning of July.

Because *Mikhail Somov*, with its now reduced complement, is well provisioned for another 10 months, evacuation plans are, so far, of a contingency nature only. A specially adapted Mi-8 helicopter is on board the *Vladivostok* to evacuate the *Somov* personnel in case of emergency.

For the moment, *Vladivostok's* chief role will be to try to break a channel through to the trapped vessel and, if necessary, tow it out. If *Somov* could break through into the relatively young ice of the Ross Sea, the most difficult part of the mission would be over. Meanwhile, scientists are continuing their work aboard the trapped vessel, including, presumably, the seismo-acoustic probing of the sea-bed that was a major feature of the ship's original programme. It is unfortunate that Soviet plans to have all seven permanent Antarctic stations equipped with Inmarsat receiving and transmitting equipment have not yet been implemented. So far only one such station, the Molodezhnaya observatory in Enderby Land, is so equipped, while the two nearest to the trapped ship, Russkaya and Leningradskaya, are linked to Leningrad only by shortwave radio.

Vera Rich

Correction

There was an error in last week's story about laser fusion (*Nature* 27 June, p.706): the current budget for inertial confinement fusion is about \$140 million; the administration proposed halving this sum in the fiscal year 1986.

Eureka

Brussels hope after Milan chaos

Brussels

WITH the key issue of institutional reform referred, reluctantly by some, to an inter-governmental conference to take place later this year, the approval by heads of government of the European Economic Community (EEC) of the idea to give the Community a new technological dimension may be considered as, at least, one positive result at the end of the two-day summit in Milan. But a series of questions still remain unanswered.

EEC heads agreed that technology in the Community should be strengthened. Thus they agreed that there should be a close link between technological development and the effort to unify the internal market, that technological effort should be closely tied to common policies as on trade policy, that duplication of national efforts should be reduced and money pooled.

The Milan summit also endorsed a Community report on the subject but the heads of government gave no clues as to the form research collaboration would take and who would coordinate it.

Questions of financial support and structure will be considered at an *ad hoc* meeting on European technology by EEC research ministers and "other qualified representatives of governments" (which may imply industrial participation). Government heads asked the French (perhaps not insignificantly) to convene a meeting in Paris, probably on 17 July. These "technology assizes" will also include the new (from 1986) EEC members Spain and Portugal, as well as non-EEC countries Austria, Sweden, Switzerland and Norway to whom participation in projects will be open and who have expressed an interest. At Milan, two proposals were put forward, the Eureka project in which countries and companies may collaborate unilaterally or in groups according to the dictates of interest and investment capacity, and the Commission proposal which would like to see the Eureka project as part of a flexible structure coordinated on a Community level, with projects operating on the lines of the Joint European Torus at Culham in the United Kingdom.

The Commission proposal has the advantage of being able to fall back on existing structures, and it has the support of the smaller member states which need to share the financial burden of technological research investment. EEC research commissioner Karl-Heinz Narjes has suggested the creation of a corridor in the research budget as well as increasing the share of the budget itself.

The other possibility is that Eureka might develop separately, out of reach of Community control (which would please President Mitterrand) but which would be overseen by an intergovernmental agency

which would at the same time keep an eye on the Community programme and would thus coordinate the two. A possible candidate for the job is Viscount Etienne Davignon, ex-EEC research and industry commissioner and father of the much quoted ESPRIT programme which launched cooperation between European industry and universities. In this second case the ten Community countries would be treated as just another member state in

participating in Eureka projects.

Conceding that Eureka should develop separately would allow at least some member states to get going on technological research while the institutional debate in the EEC continues. For even if the Commission maintains that the proposal is based on existing operational rules, bringing about a "qualitative technological leap forward" will require, as Jacques Delors, president of the European Commission, insisted before the Milan summit, changes to inject dynamism into functioning of community institutions and encourage the creation of an internal market.

Anna Lubinska

Eureka

Companies ready to bid for funds

WHILE European heads of state have been debating whether and how to go ahead with Eureka, the French proposal for a joint European technology programme, electronics companies in Europe have been defining projects that their political masters might support. Last week GEC (United Kingdom), Siemens (West Germany), Philips (The Netherlands) and Thompson (France) announced that they have already set up a study group to make joint proposals to Eureka, while the Norwegian computer company Norsk Data and the French defence, space and electronics group, Matra, announced the development of a super-mini-computer, a "Crayette" (for scientific use) under the Eureka banner.

These are some of the first concrete developments to flow from the Eureka idea originally proposed by President Mitterrand of France to provide Europe with an alternative to participation in the US star wars programme. The speed with which they have been put in train would seem to indicate that Eureka has touched a receptive nerve.

According to British GEC, experience of the European Commission's ESPRIT programme for pre-competitive research in information technology, while positive, has indicated the need for more market-oriented cooperation in real product development, which might be stimulated by Eureka. ESPRIT itself has "barely started", but has shown that an "unbiased" organization like the Commission can catalyse cooperation, spanning national frontiers.

A further factor is money. While nothing has been agreed, French sources have suggested that France alone would be prepared to contribute an initial £100 million to Eureka. Development costs in electronics seem to be increasing exponentially, according to Philips (Eindhoven) directors, so that the potential availability of public pots of gold is stimulating. International cooperation in development would also ease the burden.

What products are in the offing? Derek

Roberts, technical director of GEC, appears to have in mind new transportation systems and traffic control systems, while pan-European cellular radio networks are also in the air. The French have also been pushing for the development of a large European super-computer to rival the Cray machines.

Meanwhile, Norsk Data, which first made its name by supplying the computer control systems for the super proton synchrotron at CERN, and Matra are working together on a more immediate project, to develop not a European Cray, but something that might be described as a "Crayette".

The two companies have already been cooperating on the development of a 32-bit mini-computer to supplant Digital Equipment's ubiquitous VAX and, says Matra, they have identified a market gap for desk-top vectorial machines. These, like the Cray, offer a degree of parallel processing, essentially performing instant matrix multiplications on columns of numbers ("vectors") many elements long. Such computers are proving essential in number-crunching operations as in quantum chemistry, atmospheric modelling and seismology, but they are at present very expensive, ranging from £1 million to £10 million.

Norsk Data and Matra, however, hope to come on the market (with the help of Eureka) in two years with a vectorial machine of some "hundreds of millions of operations per second" at a price of only £100,000 to £1 million, compared with the current price of a VAX of around £60,000. Matra is aiming at the market for 32-bit minis in 3-4 years time.

One Cambridge quantum chemist said last week that "there's no doubt that this is the way it'll go" and that the Matra/Norsk Data proposal sounded "extremely competitive". The present cost of vectorial computing is a real hindrance to its development, he said. Companies such as Convex of the United States would be in competition with the European group, however.

Robert Walgate

Bioethics

US Congress tries again

Washington

THE US Congress is trying to fill the gap left by the 1982 demise of the President's Commission on Bioethics. Last year's attempt failed when President Reagan vetoed the National Institutes of Health (NIH) authorization act, which included provisions for a panel to monitor human genetic engineering, protection of human research subjects and other issues in medical ethics.

Last month, the House of Representatives again passed the NIH bill proposed by Representative Henry Waxman (Democrat, California), largely unchanged from last year's version. A similar bill is pending in the Senate. But its fate must be uncertain. Both House and Senate bills retain the features its president found objectionable last year, in particular the detailed specification of each NIH institute's research mission labelled, in the veto message, as "micromanagement" by Congress. Both

bills also call for two new NIH institutes, one for arthritis and one for nursing. NIH has long opposed the creation of new institutes, which it says add to its administrative costs without augmenting or improving the research.

Even if there is a veto, the bioethics panel may survive. Senator Albert Gore (Democrat, Tennessee), a key supporter of the concept when he was in the House, has filed a separate measure much along the lines of the section of the House and Senate bills dealing with the panel. A bipartisan board would appoint a committee of biomedical researchers, physicians, ethicists and laymen to carry out studies of ethical problems in biomedicine. The committee would have no regulatory function and would operate as an agency of the legislative branch, akin to the Office of Technology Assessment and the General Accounting Office. There is a strong feeling among even supporters of human gene therapy that a high-level but nonregulatory panel would be the best way to evolve policy on its new medical technologies while disarming the more extremist critics. A less-favoured alternative is an expansion of NIH's Recombinant DNA Advisory Committee (RAC) which has begun to draft guidelines for human gene therapy experiments. But RAC, as part of an executive branch agency, also lacks the formal independence that many would like for an ethics panel.

Another role of the ethics panel would be to straighten out the muddle over fetal research. Under current rules, NIH may not support research posing "greater than minimal" or "unknown" risks to a fetus. Although the Secretary of Health and Human Services (HHS) is allowed to grant a waiver if an experiment has passed the scrutiny of the local Institutional Review Board, the advisory council of the relevant NIH institute and a specially empanelled Ethical Advisory Board, the Secretary of HHS has never been willing to appoint an Ethical Advisory Board. In effect, there is a total moratorium.

Under the House and Senate NIH bills, the ethics panel would be asked to study the question of experiments involving unknown or greater than minimal risks. Waxman said he had reluctantly agreed to go along with the Senate's insistence on a 3-year moratorium even on waivers in the hope that this would allow time while "tempers cool and rhetoric dies down". He noted, for example, that a study on the dangers of vaccinating pregnant women against rubella, which was carried out in 1969, before the current rules became effective, would probably now require a waiver. In that study, women planning to have therapeutic abortions were vaccinated; after the abortions took place

11 to 30 days later, fetal tissue was examined to determine if the vaccine virus was able to cross the placenta and infect the fetus. The answer, contrary to expectations from animal studies, was yes; and physicians were ordered not to administer the vaccine to pregnant women.

As a practical matter, however, the 3-year moratorium will not affect research, and Waxman said he hoped that after a "full and even-handed" review by the ethics panel, the Secretary of HHS would "proceed with the responsibilities to pursue research ethically and without regard to purely political pressures".

Stephen Budiansky

Ethics in science

Washington

ONE of tenets of the scientific method is that researchers are supposed to share their data with others who wish to verify or expand on their work. That that apparently does not always happen is the unstated but clear implication of a new report from the National Academy of Sciences* urging the adoption of guidelines by grant-making agencies and scientific journals that will normally require data-sharing.

The report recites the usual reasons for sharing data: encouraging open scientific inquiry; allowing for verification, refutation or refinement of original results; bringing multiple perspectives to bear; encouraging interdisciplinary use of data; and detecting the rare cases of fraud.

But, the report notes, researchers may be reluctant to hand over their data to other researchers for various reasons.

The report recommends that researchers make their data available by the time of publication of the initial major results: "the practice of withholding data until all possible analyses are exhausted is unnecessarily restrictive and too self-serving to advance science", it said. Those who request data should bear any incremental costs, however.

Government funding agencies, the report says, should require applicants to guarantee data-sharing or to justify explicitly, in their grant proposals, why they should be exempt from the requirement. The National Science Foundation (NSF), for example, already requires that "data banks or software" produced with NSF grant funds be made available to others; the report notes, however, that the policy could be strengthened.

Finally, the report recommends that scientific journals should require authors to provide peer reviewers with access to their original data and "strongly encourage" them to provide similar access to others, and should give more space to reports of secondary analyses and replications to encourage the process. Stephen Budiansky

*Sharing Research Data, Committee on National Statistics; National Academy Press, 1985.

Paris scents triumph

Paris

IN France, the feeling over Saturday's summit debacle over the European Community's Treaty of Rome appears to be one of sweet revenge over that De Gaulle of an Englishwoman, Mrs Margaret Thatcher, and for the years of confrontation over the British budget in Brussels. But the sweetness will certainly turn sour if the result is to damage the French-inspired Eureka project for European technology.

So far, however, all seems well. The summit has produced an agreement to hold here in Paris in the second half of this month a "technology conference" of European foreign and research ministers which should, at last, result in the definition of exactly what Eureka will be, who will run it and who is prepared to pay. But the fear among those closest to Eureka in Paris may be that the conference will be compromised by the arguments among their political masters in Europe over other issues.

The French sources were at pains on Monday to distance Eureka from the battle. "Mrs Thatcher showed herself quite constructive on technology", a French research minister spokesman said, and the split over veto rights and the Treaty of Rome was "quite another question".

The apparent conflict in Milan between François Mitterrand and the president of the European Commission, Jacques Delors, was also being played down in Paris this week. The view here is that the European Commission can be "closely involved" and that "Eureka is not in conflict with the Commission". The communique from Milan "took note of and endorsed" Delors' plan. Robert Walgate

A capital Universe?

SIR — On one side of the Atlantic, authors and editors continue to use the word universe, whereas on the other they have changed over to the word Universe. Last year, while hastily proofreading an article written for a British journal, I failed to notice that the copy editor had everywhere altered the common noun universe into the proper noun Universe. When the article was published, I realized too late that this apparently harmless alteration had changed in a significant way my intended meaning. By no stretch of the imagination can a universe become the Universe.

Elsewhere¹ I have pointed out that current knowledge and usage assign different meanings to universe and Universe. The Universe means everything, including ourselves, and no doubt our descendants in the distant future will still be trying to understand the ultimate nature of reality. Nobody knows what is the Universe. Illogically, as a proper noun in common use, Universe denotes nothing more than a model of the Universe. When editors delete "universe", they should substitute not "Universe," but "model of the Universe".

The more modest and flexible word universe usually denotes "a model of the Universe". By using universe we avoid the meaningless Universe and the terminological rigmarole of model of the Universe. Every society has its universe or model of the Universe. We in the twentieth century have our physical universe, which is still a long way from being the Universe. We debate the rival merits of big-bang and steady-state universes, and under the rubric early universe discuss the virtues of the inflationary universe. (What a mess the last sentence becomes if Universe is substituted for universe!) Our descendants, though part of the same Universe, will live in different universes. In the history of cosmology we study the rise and fall of universes, or cosmic belief-systems, or models of the Universe, but not of Universes.

Indiscriminate use of Universe confuses the real thing with a model. As a result we forget that we have precious little knowledge of the true nature of the Universe. Like our forebears in past millennia, we tend to think that the end of the search for all knowledge looms in sight. Most people in the past, in the Middle Ages for example, were convinced that they had discovered the real thing. People with different models often finished up at the stake.

There exists only one Earth, and similarly only one Universe. But the difference is vast in more senses than one. With the first we know more or less what we are talking about, but with the second we have no idea. The curious words earth and Universe found in many journals and books look most odd. Logic dictates that we capitalize proper nouns such as Earth,

Sun, Solar System, Galaxy, Local Group, and leave it at that.

EDWARD HARRISON

Department of Physics
and Astronomy,
University of Massachusetts,
Amherst, Massachusetts 01003, USA.

1. Harrison, E.R. *Cosmology: The Science of the Universe* (Cambridge University Press, New York, 1981).

Imunovir

SIR — Replying to my letter¹ criticizing the UK campaign for the launching of isoprinosine (or Imunovir) as an effective anti-herpes treatment, Helen J. Wright made, on behalf of the laboratory commercializing the drug, numerous assertions and insinuations². I feel obliged to comment very briefly on some of them; the readers of *Nature* will have no difficulty in seeing most of the contradictions and the reasons for this defence.

To call W.H. Wickett's drugstore book³ an objective "independent text" on herpes does not correspond to reality. The author conducted, on behalf of the manufacturers (Newport Pharmaceuticals), a non-controlled clinical trial on herpes patients, and he reported his results in this book. If the name of the "test drug" is not mentioned until p.217, and this on the express request of the manufacturers, it is mentioned 12 times in the references between pages 218 and 225. The tantalized reader will have no difficulty in discovering the substance producing the wonderful results. But there is something more puzzling: Wickett apparently published his results only in this book and reported them at a couple of scientific meetings. Why do such extraordinary results, and those obtained by others using the same drug and cited by H. Wright, remain at the stage of meeting abstracts, which even a computer survey fails to pick up?

Bona fide readers who saw the various assertions concerning the drug are still wondering what the drug really does for herpes patients, whether it cures the "bouts"^{2,3} or diminishes to "nearly zero" the frequency of the attacks⁴. If this were the case, the number of patients should indeed decrease.

I am not surprised that articles such as those in the *Financial Times* and the *Sunday Times* promising a miraculous cure for diseases of great psychological impact — even if the titles of articles are written, as H. Wright suggests, by "irresponsible sub-editors" — are found "well balanced" by those whose main concern is the sale of the drug. I maintain that articles written for the lay public should be underplayed rather than exploit the spectacular. Readers of lay publications do not usually read journals such as *Nature*, where statements are made with extreme care, despite the fact that, or perhaps because, the readers

take everything with several pinches of salt.

This is the case of C. Wenz's article where the author cautions, because of the placebo effect, against drawing hasty conclusions in assessing the efficacy of Imunovir "until more clinical trials have been completed"⁴.

H.J. Wright has confirmed my belief that only those with vested interests would reject my criticisms and the idea of exercising self-censorship against aggressive publicity. I am certain that most will refrain from such practices and I also hope that the lay media will adopt a more restrictive policy in printing news related to health matters.

DIMITRI VIZA

Faculté de Médecine,
15, rue de l'Ecole de Médecine,
75006 Paris, France

1. Viza, D. *Nature* 313, 344 (1985).
2. Wright, H.J. *Nature* 314, 126 (1985).
3. Wickett, W.H. Jr *Herpes: Cause and Control* (Pinnacle, New York, 1982).
4. Wenz, C. *Nature* 311, 404 (1984).

Fuels from the farm

SIR — There seems to be an error in Anna Lubinska's article (*Nature* 4 April, p.395). It is stated that 8 million tonnes of cereal surpluses could produce 2 million tonnes of ethanol and that 1 hectare of land can produce 4 tonnes of ethanol. This would mean that 1 hectare of land produces 16 tonnes of cereal grain, but that is surely not the case. A good average yield through the countries of the European Communities is nearer half this figure. It should be borne in mind that the vagaries of the cereal intervention system ensure that it is the less efficient farmer who actually produces the surplus. These farmers are more likely to produce 4 tonnes per hectare of low quality cereals. The immense complications of agricultural production should be borne in mind when calculating potential hydrocarbon energy production by farms.

P.J. RYAN

91 Elliman Avenue,
Slough, Berks SL2 5BD, UK

Nullius in verba

SIR — It was good to have the exact meaning of the Royal Society's *Nullius in verba* explained by the president (see *Nature* 23 May, p.272). In all fairness to young scientists, however, he should have urged caution in taking the society's motto as a rule to walk by. For those who reveal too early their realization, which will come soon enough if they are independent thinkers, that many authorities of the day are emperors with very threadbare clothes, the going is likely to be rough.

P. R. BELL

Department of Botany
and Microbiology,
University College London,
Gower Street,
London WC1E 6BT, UK

Uniting mechanics and statistics

An adventurous scheme which seeks to incorporate thermodynamics into the quantum laws of motion may end arguments about the arrow of time — but only if it works.

THE logical relationship between the laws of mechanics and those of thermodynamics deserves more attention than it usually receives. Thermodynamics and statistical mechanics are ways of describing the behaviour of macroscopic systems made from components whose behaviour is determined by the laws of mechanics, classically those of Newton (as amended), but otherwise the equations of motion of quantum mechanics. Where the first law of thermodynamics is concerned, there is no difficulty. In both classical and quantum mechanics, total energy is a constant of the motion and is thus always conserved, at least in a closed system.

The difficulty arises chiefly with the second law of thermodynamics, and not only because there is such a variety of ways in which this principle can be defined. But now a group of three theoreticians has put forward an intriguing way in which the laws of quantum mechanics may be modified so as to incorporate the second law from what appears to be the outset (Beretta, C. P., Gyptopoulos, E. P. & Park, J. L., *Il Nuovo Cimento B* **87**, 77–97; 1985). Whether the modification proposed is sufficient, only time will tell, but the objective seems well worth the trouble Beretta *et al.* have taken.

The difficulty is well illustrated by the way in which some kind of correspondence is established between the mechanical behaviour of a system and its thermodynamic properties. For more than a century, people have been brooding on the paradox that while the laws of classical (and, for that matter, quantum) mechanics are symmetrical with respect to time inversion, the second law selects from all possible trajectories of motion only those corresponding to a continual increase of the entropy. The arrow of time is conjured like a rabbit from a hat.

The definition of entropy in terms of the mechanical properties of the constituents of a system is similarly clouded. The classical model is Boltzmann's *H*-theorem (1872), which shows that the rate of change with time of a certain mathematical construct from the probability distribution of single particles in phase space will always be zero or negative. So Boltzmann argued, his quantity *H* is admirably suited to be the negative of what is known in thermodynamics as entropy. This is argument by analogy, but none the worse for that — if it works.

Since Boltzmann's time, there has

accumulated a rich literature on the implied paradox of the conflict between the irreversibility of macroscopic processes and the reversibility (in time) of the laws of mechanics and thus of microscopic processes. Indeed, the argument was begun by Loschmidt in 1876, but now even elementary text-books of thermodynamics reckon to give some kind of account of it.

The standard explanation is that the apparent paradox is not a paradox at all, but a confusion about timescales. Any measure of entropy, that derived from Boltzmann's *H* or otherwise, will fluctuate (and so decrease as well as increase on a short timescale), which is not inconsistent with the notion that the average value of the entropy should increase steadily over long periods of time (or remain unchanged when the system is in equilibrium).

Much the same is said of the recurrence paradox, based on the observation due to Poincaré that the point in phase space (momentum as well as position) representing the state of a classical system will return to more or less the same place after a sufficient length of time. On the face of things, that means that non-equilibrium states of a system will repeatedly recur. The standard resolution of that paradox is the observation that, for any realistic system, the interval of time between recurrences will be huge, much greater than, say, the age of the Universe. Again there is nothing wrong with these arguments, but they are far from being rigorous.

So why not take the bull by the horns, and build irreversibility into the laws of mechanics? That is the point from which Beretta *et al.* start. Properly, they acknowledge that they are not the first to tread this path. They work with quantum statistical mechanics, where the formalism is easier. They start from the equation of motion for the operator representing the state of a physical system, say *m*, which is, in operator language, $\frac{dm}{dt} = -i/\hbar [H, m]$, where *t* is time, *H* the Hamiltonian operator of the system and *i* and $\hbar$ the square route of minus one and Planck's constant (divided by 2π) respectively. The quantity in square brackets is the commutator of its two components, $mH - Hm$.

The natural way to proceed is to assume that this equation is modified in such a way that the right-hand side is some other function of the state operator *m* than in the standard form. The objective is to find a form of the function which is compatible

both with what is known of the evolution of thermodynamic systems and, perhaps more important, the dynamics of real microscopic systems. Beretta *et al.* have convinced themselves that the function they are seeking cannot be a linear function of *m*. What they propose is the addition to the right-hand side of the quantum equation of motion of a particular function of *m* which, by including both the square root and the logarithm of the state operator of the system, is non-linear enough to satisfy anybody's taste.

Almost magically, the system has some of the obviously necessary properties. For example, for a system in a pure quantum state, say that represented by a solution of Schrödinger's equation, the extra terms vanish and the simple form of the equation of motion applies. Similarly, constants of the motion in the new system are also constants of the motion determined by the simpler equation of motion.

What can be said about the entropy? In reality, the state operator *m* is the equivalent of what is called the density matrix in quantum statistical mechanics, which is why Beretta *et al.* define entropy in terms of the operator $m \log m$, where the logarithm is the natural logarithm of the operator *m*. Specifically, the entropy of the negative of the trace of this operator multiplied by Boltzmann's constant; the authors are able to show that it increases (or does not increase) in the course of time.

So is this a demonstration that the laws of mechanics and of thermodynamics can indeed be combined? Not quite. For one thing, there are various mathematical problems that make some of the steps in the argument conjectural. Worse still, some of the operator functions in the formalism are sometimes undefined. But the system does have the merit of hanging together — the paper now published extends to composite systems the treatment of one-component systems published a year ago.

None of this implies that the arguments about the reconciliation between microscopic reversibility and macroscopic irreversibility will now be stilled. Indeed, while for as long as the present justification of the basis of statistical mechanics holds water, there will be many who say that what Beretta *et al.* have done is strictly unnecessary. But this is a field in which the proof of the pudding is in the eating.

John Maddox

Acquired immunodeficiency syndrome

Laying the groundwork for neutralizing AIDS-linked virus

from Jerome E. Groopman

WITH what success does the human immune system produce antibodies that neutralize the human T-lymphotropic virus type III (HTLV-III) — the virus that is aetiologically linked to the acquired immunodeficiency syndrome (AIDS)? And can the neutralizing antibodies be exploited for prophylaxis or therapy of AIDS? Two papers elsewhere in this issue^{1,2} report that such antibodies are present in some sera both from AIDS patients and from people who have been exposed to the virus but are asymptomatic. The relatively low titres of neutralizing antibodies in HTLV-III-infected persons should not dissuade investigators from pursuing their clinical relevance; the feline leukaemia virus often elicits low titres of neutralizing antibodies *in vivo*,³ yet vaccines of some effectiveness, which exploit at least in part this antibody response, have recently been developed.

Several studies that would help to define the significance of HTLV-III neutralizing antibodies *in vivo* immediately suggest themselves. Comparison of the titres and types of these antibodies in asymptomatic and AIDS patients can readily be accomplished. Particularly interesting would be longitudinal studies to examine whether the neutralizing antibodies persist in the sera of those who remain relatively healthy despite exposure to the virus but diminish in those to progress to AIDS. Of course, the clinical situation may not be so simple. Like visna virus, to which it bears some similarity, HTLV-III may alter its surface antigenic determinants over time and 'escape' from the protective effects of neutralizing antibodies^{4,5}. Serial culture of HTLV-III with testing of neutralizing antibodies on the host's own isolates is one approach to determine if *in vivo* there are selection pressures that favour the emergence of antigenically altered virus. Initial reports of apparently healthy adults in Central Africa who are both antibody positive and virus positive for HTLV-III should make this population particularly valuable in addressing the clinical role of neutralizing antibody⁶. Because antigenic determinants on the gp120 envelope protein of the gp41 transmembrane protein of the virus^{7,9} are likely to be the primary sites of interaction with the putative cellular receptor for HTLV-III^{10,11}, it is to be predicted that neutralizing antibodies are directed against gp120 and/or gp41 determinants. The wide antigenic variability in envelope-related antigens from different HTLV-III isolates¹² makes testing the ability of geographically-distant sera to neutralize geo-

graphically-distant isolates a valuable means of identifying conserved epitopes for neutralizing antibodies.

Recently, a retrovirus that cross reacts in serological tests with HTLV-III has been isolated from African Green monkeys¹³. Studies on the neutralization of HTLV-III by sera from these monkeys and of the ability of human HTLV-III neutralizing antibodies to recognize the surface determinants of the monkey virus may also help to localize highly conserved epitopes that would interact with cell-surface receptors for the viruses.

HTLV-III, LAV and ARV — all clearly variants of one type of human retrovirus — have been cloned and sequenced, thereby providing ready access to recombinant (made by recombinant DNA technology) viral peptides^{14,17}. Peptides of the envelope protein can be used *in vitro* to absorb out the neutralizing activity and localize the epitopes that elicit antibodies. It is of interest to determine whether some neutralizing antibodies might react with the putative gene products (termed sor, P', Q and 3'ORF, E', F) that may have functional roles, particularly in the activation of cellular genes (trans-activation). The availability of these products will allow the reverse test: do polyclonal or monoclonal antibodies raised in animals against any of these gene products have *in vitro* neutralizing capacity? To obtain the equivalent human antibodies, it may be possible to use B-cells from people with neutralizing antibodies.

Ultimately, *in vitro* studies will be needed to determine the clinical signifi-

cance of neutralizing antibodies for either prophylaxis or therapy of HTLV-III infection. Initial tests of prophylaxis with hyperimmune globulin and of therapy with pools of high-titre neutralizing antibodies can be undertaken in primates which are susceptible to infection with HTLV-III. Simple vaccination with native or recombinant peptides that elicit neutralizing antibodies may not be sufficient to protect primates against challenge with HTLV-III, but such an approach can readily be examined. The role of the cellular immune response (in addition to the antibody response) in preventing or suppressing human retroviral infection is still unaddressed. Innoculation of primates with forms of HTLV-III that are defective in functional genes but still display structural antigens, or with recombinant constructions of HTLV-III with other viruses, such as vaccinia, should permit the study of both cellular immune and antibody responses.

The HTLV-III research community will now coordinate its clinical and scientific forces to build on the new observation of neutralizing antibodies in the hope of constructing strategies to outwit this threatening human retrovirus. □

1. Weiss, R. A. *et al.* *Nature* **316**, 69 (1985).
2. Robert-Guroff, M. *Nature* **316**, 72 (1985).
3. Hardy, W. D. *et al.* *Cancer Res.* **36**, 582 (1976).
4. Gonda, M. *et al.* *Science* **227**, 173 (1985).
5. Clements, J. E. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **77**, 4454 (1980).
6. Saxinger, W. C. *et al.* *Science* **227**, 1036 (1985).
7. Allan, J. S. *et al.* *Science* **228**, 1091 (1985).
8. Barin, F. *et al.* *Science* **228**, 1096 (1985).
9. Ruby, W. *et al.* *Science* **228**, 593 (1985).
10. Dalgleish, A. G. *et al.* *Nature* **312**, 763 (1984).
11. Klatzmann, D. *et al.* *Nature* **312**, 767 (1984).
12. Ratner, L. *et al.* *Nature* **313**, 636 (1985).
13. Kanki, P. *et al.* *Lancet* **i**, 1330 (1985).
14. Ratner, L. *et al.* *Nature* **313**, 277 (1985).
15. Muesing, M. *et al.* *Nature* **313**, 430 (1985).
16. Wain-Hobson, S. *et al.* *Cell* **40**, 9 (1985).
17. Sanchez-Pescador, R. *et al.* *Science* **227**, 484 (1985).

Jerome E. Groopman is in the Division of Hematology/Oncology, New England Deaconess Hospital, Harvard Medical School, Boston, Massachusetts 02215, USA.

Lasers

Dynamic holograms from crystals

from Malcolm Gower

WHEN certain crystals are exposed to spatially non-uniform light fields, light-induced refractive index changes occur, caused by the migration and re trapping of charges. Known as the photorefractive effect, the phenomenon was first reported nearly twenty years ago as the mechanism responsible for the self-focussing of laser light in crystals of LiNbO₃ and LiTaO₃, and has since been observed for many more crystals, including CdS, InP and GaAs. It was soon recognized that photorefractive crystals used as three dimensional media for holographic recording (volume holograms) could contain extremely high densities of information,

up to 10¹² bits per cm³, yet retain exposure sensitivities comparable to those of the best photographic emulsions (at least 100 μ J per cm²). But, unlike conventional media, photorefractive crystals can be used for dynamic or 'real-time' holography in which the writing and reading processes occur virtually continuously and simultaneously. Recently there have been some remarkable discoveries in the field that may lead to a host of applications in real-time optical data processing — from optical image amplifiers, enhancers and pattern recognizers to optical switches and memories¹.

Unlike the relatively slow serial proces-

sing that occurs in conventional electronic devices, dynamic holograms allow rapid, parallel-bit processing of information. Because the image wave is a complex or phase conjugate of the subject wave, real-time holography is closely analogous to the nonlinear optical method of degenerate four-wave mixing used in producing phase conjugate (wavefront reversing) mirrors (see Fig 1a; ref.2). Thus these crystals can also be used as phase conjugate mirrors to correct distortions on optical wavefronts and to project complex images.

The hologram in photorefractive crystals results from the migrating charges, which may be either positive, as in BaTiO₃ or negative, as in Bi₁₂SiO₂₀ and Bi₁₂GeO₂₀. The charges are generated by photoexcitation of electrons or holes from impurity or dopant levels in the band gap of the crystal. The interference pattern set up between the subject and reference waves during recording causes the mobile charges to diffuse or drift (photoconduct) and become retrapped in the darker regions of the interference pattern, leaving behind trapped ionized centres (Fig.1b). This spatial separation of charge creates local piezo-electrically induced stresses and strains caused by electric fields whose gradients (by Poisson's equation) are a maximum in regions of maximum space-charge density, ρ_{sc} , so that the space-charge field E_{sc} is often phase shifted from the original interference pattern (Fig. 1b). The space-charge field then produces local variations of the refractive index,

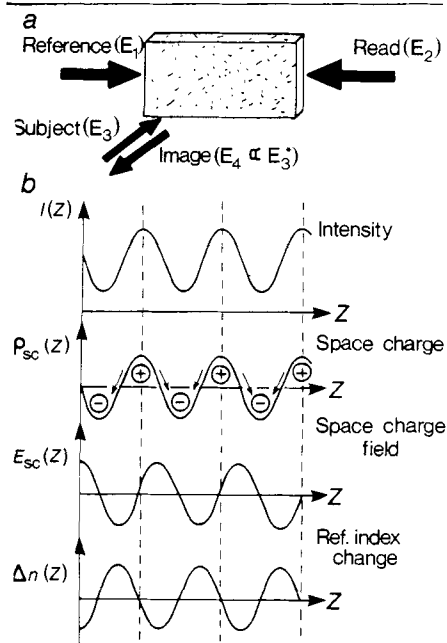


Fig.1 *a*, Dynamic holography: arrangement for producing real-time holograms and degenerate four-wave mixing phase-conjugate mirrors. *b*, Photorefraction: an interference pattern causes charges produced in bright regions to migrate into and become trapped in dark regions, creating a phase-shifted space charge field, which causes variations in refractive index. The case is shown for negative charge carriers. Z is the space coordinate; for other variables see text.

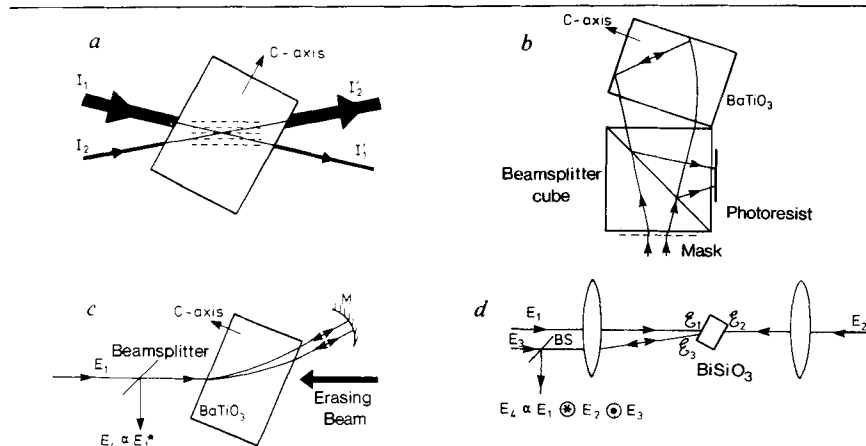


Fig.2 *a*, Two-beam coupling in phase-shifted holograms. *b*, High-resolution image projector using a self-pumped BaTiO₃ crystal. *c*, Self-pumped, optically bistable phase-conjugate reflector. *d*, Spatial convolution $\otimes$ and correlation $\odot$. $\epsilon_1, \epsilon_2, \epsilon_3$ are Fourier transforms of the writing and reading beams E_1, E_2, E_3 .

Δn , throughout the medium as a result of the linear electro-optic Pockels effect. The hologram may be permanently 'fixed', by applying a biasing voltage across the crystal during recording, or can be erased by uniform illumination of the crystal.

The maximum modulation of both the interference pattern and the induced space-charge field occurs when the recording beams have the same intensity, irrespective of their absolute values. Thus, in contrast to most other mechanisms used to produce phase-conjugate mirrors, the normalized image-wave light intensity (phase-conjugate reflectivity) produced by the reading process is insensitive to the absolute intensity of the recording beam. In BaTiO₃, for example, which has an extremely large electro-optic effect, an image-intensity gain of 100 has been observed using low-powered continuous lasers³. As the crystal behaves like a phase-conjugate mirror with gain, oscillation could be produced between it and a relatively poor, non-aligned reflector, such as a kitchen spatula. When the spatula is moved, oscillation reappears at its new position, demonstrating that the mirror can automatically track its whereabouts.

If the intensity of the recording reference beam is about half that of the subject beam, then during reading maximum reflectivity for transmitted light occurs near the edges of patterns placed in the subject beam, where the intensities of the two recording beams are equal. Such preferential enhancement of the edges of the image is often beneficial when trying to recognize complicated patterns. Because the rates for holographic recording and erasure depend on the photoconductivity of the crystal, which in turn is proportional to the intensity of illumination, using high light intensities for recording and erasure increases these rates. Thus, in BaTiO₃, for example, a hologram that takes a few seconds to record using a low-powered He-Ne laser requires less than 10 nano-

seconds when an intense (20 MW per cm²) pulsed laser is used. On the other hand, provided the crystal is kept dark, the recorded hologram can be stored for long periods, from a few hours to over 10 years depending on the crystal species. Storage is limited only by the dark conductivity of the crystal, which eventually redistributes the charge carriers.

If a weak and a strong beam intersect in a crystal (Fig. 2a), most of the energy in the strong beam can be transferred to the transmitted weak beam (since the Bragg reflection of one wave from the hologram is phase shifted by $\pm \pi/2$ from the transmitted wave of the other, the two waves will interfere constructively or destructively if the hologram is additionally shifted by $\pi/2$). Because the volume hologram written in the crystal is often phase shifted from the original optical interference pattern (Fig. 1b), the two writing beams can exchange energy by their self-diffraction from the hologram. This 'two-beam coupling' effect, although complicating the understanding of many observations in photorefractive dynamic holography, leads to some remarkable effects. For example, in BaTiO₃, where diffusion processes alone cause charge migration, the hologram phase shift is $\pi/2$ and intensity gains of 600 have been observed for the transmitted weak beam. A two-dimensional laser-beam steering device, using two-beam coupling for the potential random accessing of optical memories, has recently been constructed⁴.

Furthermore, two-beam coupling between a single laser beam and scattered light propagating in the appropriate direction in a BaTiO₃ crystal can lead to self-defocussing or 'fanning' of the beam (see cover picture). If there is total internal reflection of the fanned beam from the far corner of the crystal, phase-conjugate reflections can be produced with a reflectivity of up to 30 per cent by using just one input beam to the crystal⁵. We have recently used such a self-pumped phase-conjugate mirror, with the arrangement

shown in Fig. 2b, to project speckle-free images with submicron resolution. Such a system may be useful in photolithography, as used in the production of silicon chips, for projecting high-resolution images over a large area.

Because two-beam coupling effects are inherent in phase-shifted volume holograms, a self-starting, self-pumped phase-conjugate mirror can be produced by feeding back the transmitted fanned beam to the crystal (Fig. 2c; ref. 6). When the intensity of a separate erasure beam is continuously varied, the reflectivity of this device can have multiple values and exhibit hysteresis effects. This optical bistable behaviour may prove useful for high speed switching and memory elements in future optical computers^{7,8}.

By using lenses to focus and take Fourier transforms of the writing and reading beams, dynamic holography has been used to perform the real time operations of convolution and correlation in crystals of Bi₁₂SiO₂₀ (Fig. 2d; ref. 9). These operations are extremely useful in complex image pattern recognition, and for image deblurring and enhancement.

Furthermore, by using relatively large intensities to illuminate the subject, dynamic holography in photorefractive media can be used to create a beam that

is the mathematical division of two images¹⁰. By making use of the relatively slow recording and erasure times of these media, it is also possible to perform the real-time subtraction and differentiation of images. Compared to the numerical computational techniques now in use, optical images can in principle be manipulated much more rapidly by direct optical methods.

With their many potential applications in dynamic holography, from optical memory storage and switching to image projection and synthesis, photorefractive media may well produce a revolution no less dramatic than that caused by the discovery of photographic materials over a century ago. □

1. Günter, P. *Phys. Rep.* **93**, 199 (1982).
2. Gower, M.C. *Nature News and Views* **308**, 110 (1984).
3. Feinberg, J. & Hellwarth, R.W. *Opt. Lett.* **5**, 519 (1980).
4. Rak, D., Ledoux, I. & Huignard, J.P. *Opt. Comm.* **49**, 302 (1984).
5. Feinberg, J. *Opt. Lett.* **7**, 486 (1982).
6. Kwong, S.-K., Cronin-Golomb, M. & Yariv, A. *Appl. Phys. Lett.* **45**, 1016 (1984).
7. Abraham, E., Seaton, C.T. & Smith, S.D. *Scient. Am.* **63** (2), 85 (1983).
8. Tanguay, A. R. *Opt. Engin.* **24**, 2 (1985).
9. White, J.O. & Yariv, A. *Appl. Phys. Lett.* **37**, 5 (1980).
10. Ja, Y.H. *Opt. Comm.* **44**, 24 (1982); *Appl. Phys. B26*, 21 (1985).

Malcolm Gower is in the Laser Division, SERC Rutherford Appleton Laboratory, Chilton, Oxon OX11 0QX, UK.

Immunology

Immune response restored by gene therapy in mice

from Michael Steinmetz

Two classes of cell surface molecules encoded in the major histocompatibility complex (MHC) genes are known to guide T cells of the immune system in their recognition of foreign antigens on cell surfaces. In general, cytotoxic T cells recognize foreign antigens in association with class I molecules, whereas helper T cells co-recognize antigen and class II molecules. The precise molecular interactions that lead to the detection of foreign antigens in the context of MHC molecules (termed MHC restriction) are not known and the role that the MHC antigens play in the development and functions of antigen recognition by T cells is poorly understood. But it is beginning to look as if transgenic mice will provide some answers.

Over the past five years MHC genes, including those encoding for the T-cell restriction elements, have been cloned and sequenced both from mouse and man (see refs 1 and 2). By transfecting genetically distinct cells growing in culture with class I and II genes it has been shown that the molecules they encode function appropriately as T-cell restriction elements³. Currently, these types of gene transfections are being combined with exon-shuffling experiments to identify

which protein domains in class I and class II molecules are recognized by cytotoxic and helper T cells, respectively. One hopes that these experiments, perhaps in conjunction with X-ray crystallography, will eventually define the amino-acid residues that make contact with T-cell receptor molecules and those that may make contact with foreign antigen.

Gene transfer into somatic cells, however, is of limited use. Questions concerning the expression and regulation of genes during development and in differentiated cells, and their influence on the developing individual, are best approached through germline DNA by, in particular, micro-injecting genes into the pronuclei of fertilized eggs (to produce transgenic animals) or using retroviral vectors to transmit genes into early mammalian embryos (or bone marrow cells)⁴. With respect to the immune system there are a series of questions concerning the generation of the diverse repertoire of B-cell and T-cell antigen receptors that can be studied in transgenic animals. In the past two years, several investigators have introduced functional genes for antibody molecules into the mouse germ line to study the tissue and cell specificity of

regulation, and the mechanism of allelic exclusion, of immunoglobulin genes in B cells^{5,7}. Now, transgenic mice have been produced with cloned MHC genes to find out whether the introduced gene can be expressed in a functional way.

Frels *et al.*⁸ have introduced a porcine MHC class I gene into mouse germline DNA and have shown that the foreign gene is accurately expressed in the transgenic mouse. All cells examined display the porcine class I molecule on the surface, and skin transplanted from the transgenic mouse to a normal mouse of the same strain is rapidly rejected. It is known that genetic differences in class I molecules will induce rapid graft rejections — hence they are also termed transplantation antigens. The findings of Frels *et al.* indicate that the porcine class I molecule is recognized by mouse T cells as a transplantation antigen and that the immune system of the transgenic mouse has become tolerant to the porcine molecule. Whether the porcine class I molecule can also function as a restriction element in T-cell-mediated antigen recognition in the transgenic mouse still remains to be investigated.

A related question, namely whether class II antigens encoded by genes introduced into transgenic mice can function as T-cell restriction elements, has now been answered in elegant experiments carried out by three independent groups, two of which describe their results in this issue^{9,10} whereas the third has a paper submitted (L. A. Pinkert *et al.* personal communication). Normally, mice have two different class II molecules, A and E, both of which are composed of distant α - and β -chains. For their experiments, all three groups have made use of strains of mouse that have a 5'-deleted E_α gene and can no longer synthesize the E_α molecule. As a result, their E_β molecules, which are produced as normal, remain in the cytoplasm of the cell rather than forming a heterodimer with E_α on the cell surface. The E_α -deficiency of these mice has previously been correlated with their inability to generate an immune response against certain antigens, for instance to the synthetic copolymer composed of glutamic acid, lysine and phenylalanine (GLPhe).

To perform gene replacement therapy the different investigators used previously cloned intact E_α genes derived from two different mouse strains which contain almost identical E_α polypeptide chains. The intact E_α genes were micro-injected into fertilized eggs of the immune-deficient mice, leading to the birth of a few transgenic mice that contained several copies of the introduced E_α gene. These genes are inherited in a mendelian fashion by the offspring of the transgenic mice, indicating that they are integrated into one autosomal chromosome, and that they are transcribed into messenger RNA molecules of the right size and, more importantly, only in those cells that normally

express class II genes. In contrast to the immune-deficient mice used as recipients, the transgenic mice display an E_a/E_β complex on the appropriate cells indicating that the introduced E_a gene encodes an E_a polypeptide that associates with the host-encoded E_β molecule. Furthermore, Le Meur *et al.*⁹ and Pinkert *et al.* find that the expression of the transgenic E_a gene can be correctly induced in macrophages by γ -interferon. In contrast, Yamamura *et al.*¹⁰ find that the introduced E_a gene and the endogenous E_β gene are constitutively expressed in macrophages. The reasons for the discrepancy are not clear. Taken together, these results show that accurate regulation of the E_a gene can be achieved (despite presumably random DNA integration) with cloned DNA fragments that contain, in addition to the E_a gene, as little as 2 and 1.4 kilobases of 5'- and 3'-flanking sequence, respectively.⁹

The transgenic mice were also shown to have a restored immune response against the antigen GLPhe. Unlike the mice that did not contain an intact E_a gene they could specifically recognize the antigen as measured by T-cell proliferation and antibody production^{9,10}. This shows that reconstitution of the proper class II restriction

element, which is needed to mount an immune response against the particular antigen tested, has been achieved.

These experiments suggest that it will be possible to endow mice with new MHC restriction elements that can function properly in the generation of an immune response. Until now, manipulation of MHC genes in living animals has required the breeding of F₁ hybrid mice, genetic recombination experiments, the time consuming isolation of mutant mice and the complicated construction of chimaeric mice. Now, new, additional or *in vitro*-modified MHC genes can be introduced into transgenic mice where their effects can be studied *in vivo*. □

1. Steinmetz, M. & Hood, L. *Science* **222**, 727 (1983).
2. Steinmetz, M. *Trends biochem. Sci.* **9**, 224 (1984).
3. Steinmetz, M. in Ferrone, S. *et al.* (eds) *Human Class II Histocompatibility Antigens* (Springer, Berlin, in the press).
4. Gordon, J. W. & Ruddle, F. H. *Gene* **33**, 121 (1985).
5. Storb, U. *et al.* *Nature* **310**, 238 (1984).
6. Grosschedl, R. *et al.* *Cell* **38**, 647 (1984).
7. Rusconi, S. & Köhler, G. *Nature* **314**, 330 (1985).
8. Frels, W. I. *et al.* *Science* **228**, 577 (1985).
9. Le Meur, M. *et al.* *Nature* **316**, 38 (1985).
10. Yamamura, K. *et al.* *Nature* **316**, 67 (1985).

Michael Steinmetz is at the Basel Institute of Immunology, Grenzacherstrasse 487, CH-4005 Basel, Switzerland.

Photoelectrochemistry

Quantization effects in superlattice photoelectrodes

from Mary Archer

ARTHUR NOZIK's strained-layer superlattice photoelectrode, reported in this issue (*Nature* **316**, 51; 1985), is a neat example of technology transfer. Superlattice or multiple-quantum well structures are two-dimensional solids in which thin layers of two different materials A and B, usually semiconductors, are alternated (ABAB) or sandwiched (BAB) one with the other. Electrons localized in ultra-thin (roughly 3–25 nm) semiconducting layers produce unique solid-state physics in superlattice structures, which is now being applied to the possible production of better transistors and solid-state lasers. Nozik has used this unique physics to produce a photoelectrode that could be the prototype of a new generation of photoelectrochemical devices for the conversion of solar energy to electrical power.

What is special about the electrons trapped in a thin-layer region? As every first-year student of quantum mechanics knows, the translational energy levels E available to a particle of mass m in a one-dimensional box of length a are given by $E = n^2 h^2 / 8ma^2$, where h is the Planck constant and n (>1) is the quantum number. An electron localized in a layer of sufficiently small thickness a behaves differently from an electron free to move over larger distances because its energy levels

are coarsely spaced rather than quasi-continuous. In a superlattice structure, the energy levels in the two dimensions parallel to the layers are normal but those in the short dimension through the layers are coarsely spaced. Attractive properties of superlattice devices can be tailor-made in that the electronic and spectral properties are tailored by the thickness of the layers.

All this has been known in principle for decades, but the technology of fabrication of layered structures of good quality, essentially free from the damaging effects of surface states at the A/B interfaces, is much more recent. The production of GaAs/GaAlAs layered structures was pioneered by Bell Laboratories in the 1970s. Strain-free lattice-matched layers of these two materials can readily be produced because their lattice parameters are very similar. However, the number of pairs of materials with good lattice match is very limited. Lattice mismatch at the A/B interface cannot be tolerated because it produces a high density of surface states and superlattice structure of very poor quality. The exploitation of materials with dissimilar lattice parameters has now become possible with the discovery that suitably thin A layers can be epitaxially laid down on B to produce a strained struc-

ture, in which the lattice parameter of A is greater than in the bulk material and is matched to that of B. Fabrication, by molecular-beam epitaxy or metal-organic chemical vapour deposition of strained-layer superlattice structures with acceptably low surface-state density has been developed by the Sandia Corporation, Albuquerque in the past couple of years.

At present there seem to be three main possible applications of these quantum-well structures. The first is the production of a new generation of solid-state lasers. Existing solid-state lasers contain p - n homojunctions within a single semiconductor. Now GaAs/GaAlAs and other A/B heterojunction solid-state lasers of superior properties are being developed. These heterojunction lasers have the potential advantage that the lasing wavelength can be tailored by tailoring the thickness of the doped GaAs layers.

The second application is to the production of better field-effect transistors. Again, GaAs and GaAlAs are most often used. Doping of the active GaAs layer is achieved in a novel way by incorporating the Si dopant in the GaAlAs top layer. Electrons from the Si dopant atoms pass to the GaAs layer, rendering this layer electronically n -type but leaving it free of dopant ions. Thus, there is virtually no impurity scattering of electrons in the GaAs layer and its resistivity, which in a conventionally doped material at low and moderate temperatures is dominated by impurity scattering, is low. Consequently the losses in the transistor are less so that less electrical driving power and heat sinking are required.

The third possible application is to heterojunction photovoltaic detectors that could be fabricated on a chip and would be particularly suitable for use in combination with optical fibres. These have maximum transmission at about 0.8 eV so Si-based photovoltaic detectors,

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Superlattice structure consisting of alternating layers of GaAs_{0.65}P_{0.35} (dark layers) and GaAs_{0.75}P_{0.25} (light layers) that are 275 and 300 Å thick, respectively. The light layers form quantum wells and the dark layers potential barriers (photo courtesy of K. Jones).

with a band gap of 1.1 eV, cannot be used.

Reverting to photoelectrochemistry, quantum size effects produced by very small (<10 nm) semiconductor particles have already been reported. Nozik now describes the first quantum size effects produced in a photoelectrochemical system using a superlattice structure (see figure). The onset of photocurrent from a conventional GaAs thick-layer photoelectrode occurs at the band edge (1.42 eV) of the material, rising monotonically to a saturation value at lower wavelengths. The photocurrent action spectrum (fig. 3 of Nozik's article) of the GaAs/GaAsP strained-layer superlattice photoelectrode is quite different, showing the multiple peaks that should result from transitions between the discrete energy levels in the quantum wells.

Is this photoelectrode useful or merely interesting? It is hard to answer this question as yet. Its present quantum efficiency is poorer than that of a conventional bulk GaAs photoelectrode. On the other hand, the ability of superlattices to absorb photons of discrete energies provides a means whereby higher solar-energy con-

version efficiencies might be achieved. Conventional photoelectrodes degrade the energy of absorbed photons of energy $E > E_g$ to the band-edge energy E_g , thereby dissipating a significant fraction of solar energy as heat. This could be avoided in a superlattice photoelectrode because carrier relaxation to the semiconductor band edges should be inhibited by the presence of discrete energy levels with relatively large separation.

Two-dimensional solids are also excellent vehicles for basic physics. Their fabrication by molecular-beam epitaxy and chemical-vapour deposition is being investigated in several laboratories. "A lot of new physics will come out of two-dimensional solids" says Dr Michael Pepper, the leader of the Cavendish Laboratory's team. "For example, optical logic devices, in which thin layers are optically switched between a non-absorbing and an absorbing state, become an attractive possibility". □

Mary Archer is in the Department of Physical Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EP, UK.

Muscle structure

First sight of crossbridge crystals

from Roger Craig

TO DESCRIBE the precise atomic interactions responsible for generating the force that drives muscle contraction is the dream of muscle biologists. X-ray crystallography of contractile proteins should provide the answers but the purified proteins tend to form polymers, similar to those in muscle filaments, rather than the necessary crystals. Nevertheless, the structures of three components of the thin filaments — tropomyosin, troponin-C and actin — have all recently been resolved at atomic or near-atomic resolution (see ref. 1). And now, I. Rayment and D. M. Winkelmann have obtained and begun to analyse crystals of the head of myosin, the pivotal protein in contraction^{2,3}.

The head or subfragment-1 (S1) of the myosin molecule is the energy-transducing component of the contractile machinery and constitutes the cross-bridges, the cyclic interaction of which is responsible for the sliding of the actin filaments past the myosin filaments during force generation. There are two heads attached to one long tail in each myosin molecule; the tail is an alpha-helical coiled-coil which aggregates with other tails to form the backbone of the thick filament, whereas the heads are globular units, each comprising a heavy chain and two light chains, which together contain the sites for ATP hydrolysis and actin binding. The heads can be cleaved from the tail by proteolysis.

Myosin itself has never been crystal-

lized but much has been learnt from analysis of the purified heads and tail. Two-dimensional crystals (paracrystals) of the tail and its amino-acid sequence reveal repeating periodicities in the myosin filament backbone⁴. S1 is found to be an elongated (200 Å), curved, pear-shaped entity in the electron microscope (for example refs 5,6) although solution X-ray scattering suggests a somewhat different appearance.

But the crystallization of S1 has consistently eluded muscle researchers, leading to an assumption in some circles that the molecule could not be crystallized. Therefore, the two papers by Rayment and Winkelmann^{2,3} represent an unexpected breakthrough. Crystals of S1 diffracting to

at least 4.5 Å resolution have been grown by careful attention to the source of myosin and method of proteolysis. The relatively pure fast-twitch chicken pectoralis (breast) muscle was used as starting material, rather than the mixed-fibre muscles that are used in the more conventional rabbit myosin preparation. Hence, the myosin isozyme population was presumably more uniform than usual. Papain, in the presence of divalent cations, was used to digest the myosin instead of the more commonly used chymotrypsin. In this way S1 retains both of its light chains, known from other studies to be important in stabilizing the 'neck' region of the myosin head (chymotrypsin causes loss of one of the light chains).

It will no doubt be several years before the high-resolution structure of S1 is available. Meanwhile, in a departure from normal protocol, Winkelmann *et al.*² are studying the crystals by electron microscopy and resolving long-standing uncertainties⁵⁻⁷ about the gross structure of S1. From electron micrographs of fixed and embedded crystals sectioned in selected crystallographic planes, the authors have produced low resolution (about 30 Å) images of S1 in three different projections. Significant is the fact that optical diffraction patterns of the micrographs are similar to the X-ray patterns of the native hydrated crystals showing, reassuringly, that the micrographs present a reasonably faithful image of native S1 at 30 Å resolution. The most informative view of the crystal is that looking down the *a* axis (see figure). In this projection the long axis of the S1 molecule lies approximately in the plane of the section, and S1 molecules at different depths in the section are virtually superimposed on one another. Thus, each stain-excluding tadpole-shaped structure is effectively a projection of a single S1 molecule viewed through its thinnest dimension.

The overall shape and size of S1 can be deduced from these sections. The molecule is curved and its contour length is at least 160 Å (probably longer), providing direct and strong support for the shape and length seen in electron micrographs of shadowed myosin⁵ and in reconstructions

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Filtered image of an electron micrograph of a S1 crystal, looking down the *a* axis. The 'negative' contrast of this specimen (protein, white; solvent regions, black) is a feature of the tannic acid/glutaraldehyde procedure used to fix the crystals (courtesy of D. A. Winkelmann).

of actin 'decorated' with S1 containing both light chains⁶. It is about 60 Å at its widest, tapering distally to a less dense neck region where the light chains are located and where attachment to the tail would occur^{6,8}. This length implies that shorter estimates, based on solution-scattering data⁷, are incorrect: it seems likely that these data do not readily pick up the presence of the thin part of the myosin head.

These electron microscope studies of the crystal represent only the beginning. The authors are extending their work to a full three-dimensional reconstruction of S1 based on tilt views of the crystals. As heavy-metal derivatives and higher resolution X-ray data are obtained, the three-dimensional courses followed by the heavy and light chains should become visible. Hopefully, the structural mechanism of ATP hydrolysis will be clarified. But the culmination should be when the three-dimensional structure of S1 is brought together with that of actin: then it may be possible to decipher the nature of the

activation of the myosin ATPase by actin and possibly even to deduce the kind of structural change that occurs when these molecules interact to produce force. Meanwhile these preliminary studies both whet the appetite and strongly support the gross picture of S1 suggested by electron microscopy. They also reassure electron microscopists (and their critics) that molecular structure can be well preserved even when proteins are stained, dehydrated and observed with an electron beam under high vacuum. □

1. Schutt, C. *Nature News and Views* **315**, 15 (1985).
2. Winkelmann, D. A., Mekeel, H. & Rayment, I. *J. molec. Biol.* **181**, 487 (1985).
3. Rayment, I. & Winkelmann, D. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4378 (1984).
4. McLachlan, A. D. A. *Rev. Biophys. Bioengng.* **13**, 167 (1984).
5. Elliott, A. & Offer, G. *J. molec. Biol.* **123**, 505 (1978).
6. Vibert, P. & Craig, R. *J. molec. Biol.* **157**, 299 (1982).
7. Mendelson, R. A. & Kretschmar, K. M. *Biochemistry* **19**, 4103 (1980).
8. Flicker, P. F., Wallimann, T. & Vibert, P. *J. molec. Biol.* **169**, 723 (1983).

Roger Craig is in the Anatomy Department, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.

Crystallography

Some answers but more questions

from A. L. Mackay and P. Kramer

THERE has been an exponential explosion of activity among crystallographers following the report by D. Shechtman *et al.*¹ of a splat-cooled Al/Mn alloy showing, over extended regions, electron diffraction patterns of full three-dimensional icosahedral point symmetry. Some 25 follow-up papers, mostly preprints, are in circulation and several high-resolution micrographs of this material have been published. It remains to be seen whether the structure implies a conceptual revolution, a change of paradigm or just a shock to our complacency.

On page 50 of this issue, L. A. Bursill and Peng² publish a micrograph in which the spots fit very closely to the points in the 2-dimensional Penrose quasi-lattice (which is a section of the 3-dimensional icosahedral pattern). One characteristic feature is the arrangement of spots in a regular decagon, of diameter 14.5 Å, with an isosceles (non-equilateral) arrangement of three points inside. We might contrast this with the micrograph of the Sendai group³, where the triangles inside the rings are not resolved but which shows order over a larger area and where the long-period modulations show up clearly.

For the interpretation both of diffraction patterns and of micrographs, the problem of double diffraction is crucial⁴, as can be seen by considering five sinusoidal density waves at 72° to each other in the plane. The transform is a ring of 10 spots. If the diffracted waves are scattered again in the same specimen, the resulting diffraction pattern, the convolution of the

ring of 10 spots with itself, accounts for almost all of the observed diffraction spots (and repeated diffraction would account for the remainder). Thus, both diffraction patterns and micrographs tend to be self-similar, with ratios of the golden number, and it is difficult to decide which is the primary level.

Nevertheless, P. A. Bancel and others⁵ have obtained powder X-ray diffraction patterns using synchrotron radiation and claim to have eliminated multiple diffraction effects. Their powder pattern is indexed with six integer indices with respect to six axes (parallel to the 5-fold axes of the icosahedron). Given the earlier demonstration⁶ of how to handle standard crystallographic calculations with redundant axes such as these, the planes corresponding to the [10000] reciprocal-lattice vector turn out to have the spacing of 2.17 Å. The intercepts of this plane on the icosahedral axes, corresponding to the edges of the Penrose tiles, are $\sqrt{5}$ times as great, namely 4.85 Å, which fits reasonably with a 14.5 Å diameter for rhombic triacontahedra and would give about 5–6 atoms per acute rhombohedron. Portier and others report an edge-length of 4.60 Å for the proposed Penrose rhombohedron.

The relative stability of quasi-lattice and of normal crystalline phases has been investigated both experimentally and theoretically. G. Parthasarathy and the group in Bangalore report⁷ that high pressure (5–9 GPa, varying with composition) irreversibly converts the icosahedral phase to the crystalline, indicating that the

former is metastable. The electrical conductivity increases sharply on compression. The existence of phases ranging from fluid to crystalline has been explored by G. S. Emch and others⁸ in terms of algebraic quantum theory; it would be of interest to see applications of this theory to quasi-crystalline phases. Bak⁹ has applied the Landau theory of phase transitions to the icosahedral structures.

One of the characteristics of these icosahedral structures is now recognized to be a conflict, of a kind foreseen by, for example, A. Janner¹⁰ and D. R. Nelson¹¹, between the requirements of local and long-range ordering. (These ideas go back to those of G. Toulouse *et al.* on the topology of glasses.)

Questions of crystallography in two and three-dimensional non-Euclidean manifolds now arise. We have already seen beautiful examples of crystallography in two-dimensional manifolds in the series of papers by S. Andersson and colleagues¹² on periodic minimal surfaces; as there is a third dimension of the world of ordinary experience into which planar patterns may be curved, these provide a way of acquiring experience before trying to make sense of three-dimensional non-euclidean manifolds. But J. F. Sadoc and R. Mosseri¹³ suggest that there is an ideal structure, curved into a fourth dimension which, by disclinations, can be flattened back into ordinary space to give an icosahedral phase. This idea has still to be examined carefully, but we may well be seeing the emergence of 'non-euclidean crystallography'.

Another approach to the interpretation of the experimental results is in terms of projections onto three dimensions of six-dimensional regular crystals. N. G. De Bruijn¹⁴ (for two dimensions) and P. Kramer¹⁵ (for three dimensions) before Shechtman's discovery; and V. Elser¹⁶, and R. K. P. Zia and W. J. Dallas¹⁷ afterwards, have derived the Penrose pattern (a non-periodic tessellation of space by acute and obtuse rhombohedra of two kinds) by projection of a primitive lattice in six dimensions. The resulting pattern was called¹⁸ a 'quasi-lattice' because any vertex can be given integral indices with respect to six axes (three redundant) and corresponds to Frank's consideration of the usual hexagonal lattice (with four axes, one redundant) as the projection of a four-dimensional lattice¹⁹. The pattern is characterized by the algebraic integers which are the golden number, $\frac{1}{2}(1 + \sqrt{5})$, and its reciprocal. As the real space pattern thus involves two irrationally related steps in the same direction, the diffraction pattern resembles the convolution of these two lattices. In the light of Coxeter's discussion of zonohedra as shadows of higher-dimensional polytopes²⁰, are we back with Plato²¹ looking at shadows of reality cast on the wall of our cave?

In the icosahedral structures being

investigated spatial ordering is imperfect, but ordering of the directions of nearest-neighbour bonds persists over long distances. The three-dimensional Penrose pattern can be converted into a nematic texture by leading threads across each tile from face to face. There are 15 directions of these threads, corresponding to the 30 faces of the rhombic dodecahedron which is the basic element of the geometry of the pattern, and which connect pairs of opposite faces. Six layers each containing five sets of threads would be strongly expressed²². The construction resembles the rod packing described by M. O'Keefe and S. Anderson²³ where rods run in the four [111] directions of a cubic structure. In the icosahedral nematic texture, icosahedra connected by edges might represent thread directions²⁴; dislocations would involve line defects²⁵.

So we now have on the cards a number of theories and partial explanations of the icosahedral structures (though D. Levine and P. J. Steinhardt²⁶ suggest that other symmetries may also be possible) of which the following ingredients may have to be dissected and combined:

- Multiple twinning²⁷.
- Hierarchical packing of icosahedra³.
- Bond-orientational order²⁸.
- Three-dimensional packings of the Penrose type, probably with disordering^{16,26}.
- Icosahedral nematic ordering²⁴.
- Hypergrid descriptions and bounded projections of *n*-dimensional lattices (refs 14–16, 29, 30).
- Incommensurate (modulated) structures¹⁰.
- The flattening into ordinary euclidean space of structures which appear more perfect in a curved metric¹³.

A whole string of questions has arisen and will have to be answered before the observations now to hand are understood. What is the ideal complete ordering of the icosahedral structure, and is it really the three-dimensional analogue of the Penrose pattern? What is the actual quasi-unit-cell edge? What is the nature of the disordering from this ideal structure? And what does it mean physically that a three-dimensional structure should be the (bounded) projection of a structure of higher dimensionality, or even the result of the flattening of a structure curved out of normal space?

Further ahead is another set of questions. What are the requirements leading to the formation of icosahedral phases? Are there others with similar kinds of order? Already there are reports of degrees of order for the Al/Mn material intermediate between the icosahedral and the crystalline. How can the structure best be handled mathematically, in particular for the calculation of structure factors for comparison with experiment after allowance for double-diffraction effects? And what are the physical properties we might

expect from materials such as these?

Let us enjoy a wide-ranging bout of speculation before the definitive answers emerge from crucial observations. For it is too rarely that we see such a stimulating experiment as Shechtman's¹. □

1. Shechtman, D., Blech, I., Gratias, D. & Cahn, J. W. *Phys. Rev. Lett.* **53**, 1951 (1984).
2. Bursill, L. A. & Peng Ju Lin *Nature* **316**, 50 (1985).
3. Hiraga, K., Hirabayashi, M., Inoue, A. & Masumoto, T. *Sci. Rep. Res. Inst. Tohoku Univ.* **A32**, 309 (1985).
4. Portier, R., Shechtman, D., Gratias, D. & Cahn, J. W. (preprint).
5. Bancel, P. A., Heiney, P. A., Stephens, P. W., Goldman, A. I. & Horn, P. M. (preprint).
6. Mackay, A. L. *Acta crystallogr.* **A33**, 212 (1985).
7. Parthasarathy, G., *et al.* *Phys. Rev. Lett.* (in the press).
8. Emch, G. S., Knops, H. J. F. & Verboven, E. J. *J. Math. Phys.* **11**, 1655 (1970).
9. Bak, P. (preprint).
10. Janner, A. & Janssen, T. *Phys. Rev.* **B13**, 643 (1976).
11. Nelson, D. R. *Phys. Rev.* **B28**, 5515 (1983).
12. Andersson, S., Hyde, S. T. & von Schering, H. G. *Z. Krist.* **168**, 1 (1984).
13. Sadoc, J. F. & Mosseri, R. *J. de Physique* (in the press).
14. De Bruijn, N. G. *Proc. Konig. Ned. Akad. Wet.* (math), **A84**, 39 (1981).

15. Kramer, P. & Neri, R. *Acta crystallogr.* **A40**, 580 (1984).
16. Elser, V. (preprint).
17. Zia, R. K. P. & Dallas, W. J. *J. Phys. Math. Gen.* **A18**, 1341 (1985).
18. Mackay, A. L. *Sov. Phys. crystallogr.* **26**, 517 (1981).
19. Frank, F. C. *Acta crystallogr.* **18**, 862 (1985).
20. Coxeter, H. S. M. *Regular Polytopes* 2nd edn (Macmillan, London, 1963).
21. Plato *The Republic VII*, 515.
22. Kramer, P. *Z. Naturforsch.* **A** (in the press).
23. O'Keefe, M. & Andersson, S. *Acta crystallogr.* **A33**, 914 (1977).
24. Mackay, A. L. *Colloq. Societatis Janos Bolyai (Budapest)* (in the press).
25. Rivier, N. J. *de Physique* **43**, 293 (1982).
26. Levine, D. & Steinhardt, P. J. *Phys. Rev. Lett.* **53**, 2477 (1984).
27. Field, R. D. & Fraser, H. L. *Mat. Sci. Eng.* **68**, L17 (1984/5).
28. Steinhardt, P. J., Nelson, D. R. & Ronchetti, M. *Phys. Rev.* **B28**, 784 (1983).
29. Kramer, P. *Acta crystallogr.* **A38**, 257 (1982).
30. Kramer, P. & Neri, R. *Z. Krist.* (in the press).

A. L. Mackay is in the Department of Crystallography, Birkbeck College, London WC1E 7HX, UK; P. Kramer is Professor of Theoretical Physics, University of Tübingen, D-7400 Tübingen, FRG.

Animal behaviour

The serpent's seductive scent

from Daniel I. Rubenstein

ANIMALS are not always what they appear to be. Insects often blend into the foliage by looking like twigs or leaves; small hawkmoth caterpillars transform themselves into formidable snake-like creatures by inflating their heads to reveal large eye-like markings; and palatable butterflies, such as the vicery, can so closely resemble noxious species, such as the monarch, that they reduce their risk of being eaten without incurring the cost of acquiring and maintaining toxic substances. Clearly, deceit can help in fooling predators. Recent research, including that described on page 59 of this issue¹, shows that it can be just as useful in duping conspecific competitors.

During the past few years it has become clear that even within populations, marked differences exist in the way individuals compete for reproductive opportunities. Both theoretical^{2,3} and empirical⁴ studies have shown that the reproductive success of males adopting alternative breeding strategies can sometimes equal that of males displaying the typical form of behaviour of a species. For this to be the case, the alternative strategy, which frequently entails forgoing immediate breeding opportunities, must compensate by avoiding some of the costs and risks associated with the typical strategy. Success usually comes by copulating surreptitiously, with perhaps the best form of stealth being to adopt the guise and behaviour of females.

Such female mimicry occurs commonly in bluegill sunfish^{5,6}. In this species the dominant breeding males are large and light in colour — they build nests, drive other light-coloured males away, actively court females and provide parental care.

Other males, however, mate by parasitizing spawning pairs. These individuals are smaller, look like the more darkly coloured and vertically barred females, and simply sidle up to the spawning pair and release sperm as the female deposits her eggs in the nest of the dominant male.

Mason and Crews¹ here show that serpents have added a new seductive twist to female mimicry. Normally, mimicry enables subordinate males to hide from aggressive dominants and thus gain access to receptive females. In the garter snake, however, female-mimicking males are more than just wolves in sheep's clothing.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Male garter snakes, *Thamnophis sirtalis parietalis*, rub their chins along the backs of females during courtship. This behaviour is a specific bioassay for the snake's attractiveness pheromone. Here, three males are courting a female while a fourth has become confused and is courting a 'she-male'. Photo by D. Crews.

Not only do they increase their chances of mating with the receptive female by donning her scent, but they actually seduce, and lure, normal males away from her.

In garter snakes, courtship takes place in writhing mating swarms, or balls; each female is courted by many mates (see figure). Mason and Crews report that many of the mating swarms they studied lacked females; instead, one of the males was being courted by the others. These courted males, or 'she-males' are morphologically indistinguishable from true males, and have higher levels of testosterone and lower levels of oestrogen than true females. Yet these she-males have active female-like pheromones: the pheromone from the skin of she-males is as effective in eliciting courtship from males in experimental arenas as is the pheromone extracted from the skin of females. She-males, however, are not as sensuous as true females. When males are given a choice between a female and a she-male, only about 20 per cent of the males court the she-male. Nevertheless, she-males have a tremendous reproductive advantage. When groups of males and she-males are allowed simultaneously to court an attractive female, approximately 70 per cent of the trials end with matings by she-males. Casual observations of mating balls composed of all three classes of snakes indicated that the she-males actually seduce some males and entice them to court the she-male instead of the normal female.

So far only the seductive effect of she-males on 'he-males' has been examined. Such interactions would have predominated in natural populations when she-males first appeared and were the rarest sexual form. Given the reproductive advantage that she-males show in such interactions, the rapid proliferation of she-males seems inevitable. The effect that she-males have both on other she-males and on true females should be investigated. Do she-males interfere with the reproduction of other she-males to an even greater extent than they do with males? Do females treat she-males differently from he-males? Also, are there ecological factors that limit the viability of she-males? If behaving like a she-male truly entails no costs, however, then an evolutionary problem emerges: why aren't all males she-males? The answers to these questions will be necessary for a complete understanding of the phenomenon Mason and Crews have described. □

1. Mason, R. T. & Crews, D. *Nature* **315**, 59 (1985).
2. Dunbar, R. I. M. *Perspectives in Ethology* Vol. 5 (eds Bateson, P. P. G. & Klopfer, P. K.) (Plenum, New York, 1982).
3. Rubenstein, D. I. *Limits to Action: The Allocation of Individual Behaviour* (ed. Staddon, J. E. R.) (Academic, New York, 1980).
4. Austad, S. N. *Am. Zool. Symp.* **24** (1984).
5. Dominey, W. *Nature* **284**, 546 (1980).
6. Gross, M. R. & Charnov, E. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6937 (1980).

Daniel I. Rubenstein is Assistant Professor of Biology at Princeton University, Princeton, New Jersey 08544, USA.



This engraving of an American white pelican by John James Audubon is included in the reissue of six engravings of North American birds and in an exhibition at the British Museum of Natural History. Both mark the bicentenary of Audubon, North America's most famous naturalist. His *Birds of America*, consisting of 435 hand-coloured plates, was first published in London between 1828 and 1839. Audubon travelled over the whole United States, painting life-size pictures of native birds, many never described before. The plates were engraved on Double Elephant (991 × 673mm) copper sheets and one edition of 200 copies was printed. Most of the plates were destroyed but the American Museum of Natural History owns 12, six of which have now been restored. The money raised by selling the new issue will finance an Audubon fund at the American Museum of Natural History. The exhibition at the British Museum of Natural History, entitled "Drawn from Life", runs from 3 July to 28 September, 1985.

Stellar evolution

Brown dwarfs and hidden mass

from R. J. Tayler

IS IT possible that much of the hidden matter in the solar neighbourhood and elsewhere in the Universe is composed of brown dwarfs, low-mass subluminescent stars that do not become hot enough for nuclear fusion reactions? Until recently no brown dwarfs had been identified. Now infrared observations suggest that the star Van Biesbroeck 8B is such a brown dwarf. On page 42 of this issue L. A. Nelson, S. A. Rappaport and P. C. Joss support this identification by providing detailed calculations of the possible structure and evolution of this star.

It is generally agreed that stars whose masses are less than about a twelfth of a solar mass, the precise limit depending on the chemical composition and on some uncertainties in the physics of stellar interiors, do not attain a high enough central temperature for a useful release of nuclear energy. As a result they have only a short life as stars with significant luminosity, whereas stars whose masses are just above the limit have a very long main-sequence lifetime. The process of star formation is complex and not very well understood; it is not clear whether large numbers of stars will form below the mass limit, although some studies of the fragmentation of gas clouds have suggested that the minimum stellar mass formed might be significantly smaller. If that is true, a substantial amount of matter in the Universe could be in the form of these subluminescent stars, the brown dwarfs.

There appears to be hidden mass throughout the Universe. More than fifty

years ago, Oort pointed out that the strength of the gravitational field perpendicular to the disk of the Galaxy in the solar neighbourhood implied a local mass density greater than that provided by luminous matter. Although much interstellar gas has been discovered since then, there still appears to be hidden mass in the solar neighbourhood. In addition, the variation of rotation velocity with distance from galactic centres implies that many galaxies have higher masses than expected. Finally the dynamics of groups and clusters of galaxies indicates the existence of hidden mass on a much larger scale. It is difficult for all of this matter to be baryonic, if the hot big-bang cosmological theory is valid, but some of it certainly can be, so it is of interest to ask whether brown dwarfs are an important constituent of the Universe.

Brown dwarfs must be difficult to detect unless they are close to the Sun and previous discussions of their probable numbers has been based on estimates of the numbers of low-mass luminous stars. For masses close to and greater than that of the Sun, the number of stars in unit-mass range in a given volume of space increases rapidly with decreasing mass. This mass function must turn over at some low mass so that the total mass in the solar neighbourhood is not impossibly large, but the crucial question is whether this turnover occurs above or below the limit for nuclear reactions. Observers are not agreed on this point. G. Gilmore, N. Reid and P. Hewett (*Mon. Not. R. astr. Soc.* **213**, 257;

1985) are confident that the hidden mass cannot be in the form of dwarfs, whereas J. N. Bahcall, P. Hut and S. D. Tremaine (*Astrophys. J.* **290**, 15; 1985) believe that brown dwarfs are the best candidate.

It seems possible that some of these questions will be answered now that brown dwarfs have been observed. For some time infrared observations have provided candidate objects (see, for example, T. O. Boroson, M. S. Giampapa and J. Liebert, *Astrophys. J.* **283**, 758; 1984) but no definite identifications. Now there seems to have been a definite detection in Van Biesbroeck 8B, the companion to another very low-mass nearby star Van Biesbroeck 8A. The identification has been made using infrared speckle interferometry by D. W. McCarthy Jr, R. G. Probst and F. J. Low (*Astrophys. J.* **290**, L9-13; 1985). Now Nelson and his colleagues describe detailed models of the structure and evolution of very low mass

stars, and show that the observed luminosity and surface temperature of Van Biesbroeck 8B are consistent with its having a mass of about 0.06 solar masses and an age of about 2×10^9 years, comfortably less than the age of the Galaxy. They also suggest that the mass of Van Biesbroeck 8A is very close to the hydrogen-burning limit.

It is obviously of great interest to know that brown dwarfs do exist and to have the opportunity to check the models of low-mass degenerate stars against observations. One or two stars do not, however, account for the hidden mass, so observers will be looking for every opportunity to discover further objects to obtain estimates of their space density. As Nelson *et al.* point out, the Hubble Space Telescope and the Space Infrared Telescope Facility may provide such opportunities. □

R. J. Tayler is Director of the Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK.

Evolutionary biology

Costs of reproduction

from Linda Partridge and Paul H. Harvey

Most genetic models for the evolution of life-history patterns assume that an organism's reproduction early in life has a negative effect on its later reproduction. Therefore, reproductive rates at all ages cannot be simultaneously maximized by natural selection^{1,4}. One physiological basis for this effect may be that early reproduction diverts resources away from growth and repair, so that survival rates and fertility are subsequently lowered. Different life-history patterns can evolve as a consequence of reproductive trade-offs. For instance, if growth rates are reduced when an organism's resources are channelled into reproduction, but fertility increases with an individual's size, natural selection may favour delayed maturity and hence higher reproductive success at maturity. Conversely, if larger individuals attract the attention of predators, then early reproduction may be favoured.

These notions of reproductive costs and trade-offs provide an evolutionary explanation for the observed diversity of age-specific reproductive patterns, but until recently there has been little empirical testing of the hypothesis. Attempts to demonstrate and measure the costs of reproduction are producing what at first sight seem to be mixed results.

One major category of study examines correlations between early and late reproduction among different individuals (or clones, populations or species) living in the wild or in captivity. Both negative and positive correlations have been found. For example, the overwinter survival rates of female song sparrows (*Melospiza melodia*) that have reared broods of different sizes are positively correlated with brood size⁵.

Similarly, early and late fertility are positively correlated with survival both in and among clones of a planktonic rotifer (*Platyias patulus*)⁶. In contrast, female red deer (*Cervus elaphus*) calving in one breeding season have a reduced probability of survival to the next breeding season and, if they do survive, of then calving⁷. Only 22 of 33 such studies cited in a recent review⁸ suggest that reproductive costs do occur. But correlation is not causation, and the other studies can be interpreted on the basis of a cost of reproduction if a third variable acts simultaneously on early and late reproductive rates. For instance, individuals may differ in size, nutritional status or their fit to the habitats they occupy, particularly in the laboratory. Under such circumstances, rates of early and late reproduction may be positively correlated, because of their common correlation, with a third variable.

Therefore, when investigating costs of reproduction, it is essential to perform manipulative experiments in which other confounding variables are kept constant⁸⁻¹¹. Laboratory studies with male fruitflies (*Drosophila sp.*) provide an example of manipulations that can demonstrate costs to reproduction when the correlational evidence seems to point in the other direction. The natural correlation between longevity and age-specific reproductive activity in males is positive, because both variables are strongly correlated with body size¹². When reproductive activity is manipulated by allowing males access to an excess of sexually receptive females, rates of survival are reduced¹³. Similarly, female *Drosophila* that lay few eggs because they are kept virgin¹⁴,

or are made sterile by high temperatures¹⁵ or radiation¹⁶ have higher survival rates than females reproducing normally. Similar experiments have manipulated clutch sizes of birds in the field and hence the effort expended on parental care. One study on blue tits (*Parus caeruleus*), outstanding for its careful experimental design and large sample sizes, showed that large clutches result in weight loss and increased mortality in female but not male parents, possibly because although both sexes feed the nestlings, only the female incubates¹⁷.

Genetic manipulations of reproduction have also revealed costs. *Drosophila* females selected for high fertility late in life show increased longevity and lowered early fertility^{18,19}, and females with high fertility early in life have low longevity²⁰.

The general conclusion from properly conducted experiments is that reproduction does have costs; genetic variation has been found in the few cases where it has been looked for, thus providing the raw material for the evolution of reproductive trade-offs. Nevertheless, adequate tests of the importance of reproductive trade-offs in life-history evolution will entail experimental work under field conditions. Laboratory studies probably tend to underestimate reproductive costs, because many important features of field conditions, such as competition for food, predators and competitors are usually absent. However, making the relevant manipulations under field conditions can be extremely difficult. For instance, it is relatively easy to alter clutch and brood size in birds by addition or removal of eggs or nestlings, and hence to measure costs of parental care, but manipulation of the rate of egg laying to assess costs of egg production would be more difficult. Although there are few field studies of reproductive costs, only this type of information will allow adequate tests of some fundamental assumptions of life-history theory. □

- Williams, G. C. *Am. Nat.* **100**, 687 (1966).
- Gadgil, M. & Bossert, W. H. *Am. Nat.* **104**, 1 (1970).
- Shaffer, W. M. *Ecology* **55**, 291 (1974).
- Charlesworth, B. *Evolution in Age-Structured Populations* (Cambridge University Press, 1980).
- Smith, J. N. M. *Evolution* **35**, 1142 (1981).
- Bell, G. *Evolution* **38**, 300 (1984).
- Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. *Red Deer, Behaviour & Ecology of Two Sexes* (University of Chicago Press, 1982).
- Reznick, D. *Oikos* **44**, 257 (1985).
- Horn, H. S. & Rubenstein, D. I. in *Behavioural Ecology* (eds Krebs, J. R. & Davies, N. B.) 279 (Blackwell, 1984).
- Partridge, L. & Andrews, R. *Insect Physiol.* **31**, 393 (1985).
- Warner, R. *Evolution* **38**, 148 (1984).
- Partridge, L. & Farquhar, M. *Anim. Behav.* **31**, 871 (1983).
- Partridge, L. & Farquhar, M. *Nature* **294**, 580 (1981).
- Bilewicz, S. *Folia Biologica* **1**, 177 (1953).
- Maynard Smith, J. *J. exp. Biol.* **53**, 832 (1958).
- Lamb, M. J. *J. Insect Physiol.* **10**, 487 (1964).
- Nur, N. *Anim. Ecol.* **53**, 479 (1984).
- Rose, M. R. & Charlesworth, B. *Genetics* **97**, 187 (1981).
- Luckinbill, L. S., Arking, R., Clare, M. J., Cirocco, W. C. & Buck, S. A. *Evolution* **38**, 996 (1984).
- Rose, M. R. & Charlesworth, B. *Genetics* **97**, 173 (1981).

Linda Partridge is in the Department of Zoology, West Mains Road, Edinburgh EH9 3JT, UK; Paul Harvey is visiting the Biology Department, Princeton University, Princeton, New Jersey 08544, USA.

Is the AIDS virus recombinant?

SIR — The retrovirus which is the causative agent of acquired immune deficiency syndrome (AIDS) has recently been shown to be distantly related to the C-type retrovirus^{1,4} and rather share nucleotide sequence homology with the visna virus, a slowly replicating and pathogenic but non-oncogenic retrovirus of sheep that is a member of the Lentivirinae subfamily⁵. The phylogenetic tree inferred from the amino-acid sequences of the DNA endonuclease domain of the retroviral *pol* gene product showed that the AIDS virus is evenly separated from human T-cell lymphotropic virus type-I (HTLV-I) and Rous sarcoma virus (RSV)⁶ (our phylogenetic tree inferred from the same domain differs slightly from that of Wain-Hobson *et al.*⁶; the most distant relative is Moloney murine leukaemia virus (M-MuLV) rather than the AIDS virus). Essentially similar phylogenetic trees were also obtained from the comparisons of the reverse transcriptase domain and highly conserved regions of *gag* gene product, although they differ slightly from each other in tree topology.

In the light of these phylogenies inferred from the *pol* and *gag* sequences, the *env* gene products of the AIDS virus and murine mammary tumour virus (MMTV), a B-type retrovirus, show unexpected homology patterns. The *env* gene product comprises an N-terminal highly divergent region encoding an envelope glycoprotein (gp70 in M-MuLV) and a C-terminal conserved region coding for p15(E), known to be membrane-associated and anchor gp70 at the virion surface^{7,8}. A marked homology is seen in the membrane-associated region between C-type retroviruses (in Fig. 1a, only the homology matrix⁹ of pair M-MuLV/HTLV-I was shown). Although MMTV bears close resemblance in endonuclease sequence with RSV¹⁰, no obvious homology has been detected in the

Fig. 1 Homology matrix comparisons of amino-acid sequences of *env* gene products of four retroviruses. 1, M-MuLV¹¹ (ordinate) against HTLV-I¹² (abscissa); 2, M-MuLV against MMTV¹³; 3, HTLV-III¹ against HTLV-I; 4, HTLV-III against MMTV. The homology matrices were calculated by the method described previously⁹. Each diagonal line indicates a segment 25 residues long that shows similarity above a threshold level which corresponds to the degree of similarity that would occur by chance with a probability of less than 1.0×10^{-3} . Arrows indicate regions where significant sequence similarities were observed; per cent homology and the probability of occurrence by chance are respectively 40% and $<10^{-9}$ for 1; 37% and 7×10^{-9} for 3; 37% and 4×10^{-7} for 4. No significant similarity was observed for 2. Since HTLV-III exhibits a mosaic pattern of homology with HTLV-I and MMTV, the *env* gene products were tentatively divided into two portions.

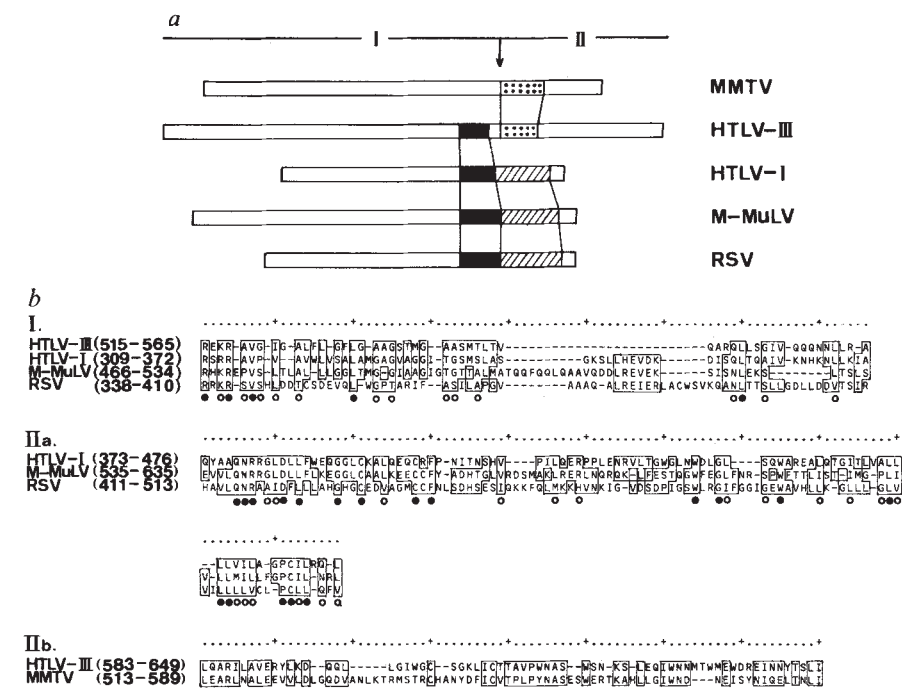


Fig. 2. Alignment of amino-acid sequences of *env* gene products among distantly related retroviruses. *a*, Homology maps along the sequence of *env* gene product. Homologous segments were indicated by vertical lines as well as filled boxes, dotted boxes or hatched boxes. The *env* gene products were tentatively divided into two regions, I and II, as in Fig. 1. *b*, Alignment. For each of the homologous segments shared by different viruses as shown in *a*, their amino-acid sequences were aligned as described previously⁹. Boxed residues indicate identities or favoured amino-acid substitutions among more than half of the viruses. —, Gap; ● and ○, positions that are occupied by identical amino acids and identical or chemically similar amino acids among all the viruses compared, respectively.

env region (Fig. 1b), suggesting that MMTV is a hybrid. Comparisons of the AIDS virus (HTLV-III) with C-type viruses and MMTV revealed unexpected homologies in the membrane-associated region. Evidently homology of the AIDS virus with other viruses drastically changes at a point around amino-acid positions 566–582 (HTLV-III¹ of the *env* gene product. In the regions before this point, the AIDS virus exhibits homology with C-type viruses, but not with MMTV (Fig. 1c). In contrast the AIDS virus shares homology with only MMTV

beyond this point (Fig. 1d). These results are summarized in Fig. 2a and alignments of the homologous regions are shown in Fig. 2b. This mosaic structure of the AIDS virus strongly suggests that genetic recombination occurred between a member of C-type virus and a B-type virus related to MMTV during evolution, generating a novel type of retrovirus, the AIDS virus.

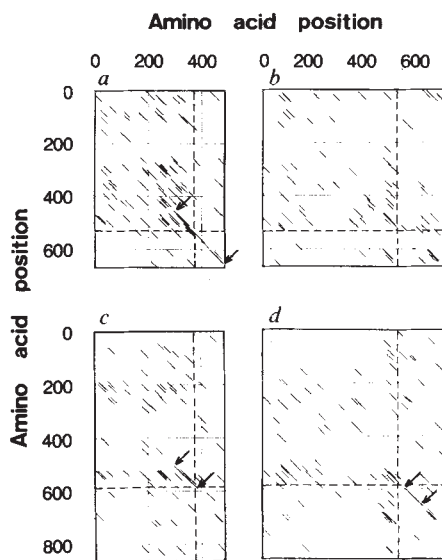
Our proposal that the AIDS virus is a recombinant of B- and C-type retroviruses is consistent with evidence^{2,14} that both the AIDS virus and MMTV use tRNA_{Lys} as a primer for (–) strand synthesis instead of tRNA^{Pro} used by other mammalian retroviruses, and that they contain very similar polypurine tracts. To confirm further, we searched amino-acid sequence similarities between the AIDS virus and MMTV beyond the membrane-associated region. Somewhat less extensive but significant homology has been detected (positions 712–742 of HTLV-III *env*-*lor*¹ and 239–269 of MMTV LTR ORF¹³. Their amino-acid sequences

VNVRQGSPLSFQTHLPGRGDRPEGIEE (HTLV-III)

VNGYKVLVYRSLPFRERLARARPFWCMSQEE (MMTV)

share 32% homology with the probability of occurrence by chance of less than 2.2×10^{-5} .

It has recently been demonstrated that sequence homology and morphologic similarity exist between the AIDS virus and visna virus⁵. Direct sequence comparison of these viruses together with



MMTV will be of particular interest in relation to our present proposal.

HIROYUKI TOH
TAKASHI MIYATA

Department of Biology,
Faculty of Science,
Kyushu University,
Fukuoka 812, Japan

1. Ratner, L. *et al.* *Nature* **313**, 277-284 (1985).
2. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. *Cell* **40**, 9-17 (1985).
3. Muesing, M. A. *et al.* *Nature* **313**, 450-458 (1985).
4. Sanchez-Pescador, R. *et al.* *Science* **227**, 484-492 (1985).
5. Gonda, M. A. *et al.* *Science* **227**, 173-177 (1985).
6. Wain-Hobson, S., Alizon, M. & Montagnier, L. *Nature* **313**, 743 (1985).
7. Dickson, C., Eisenman, R. Fan, H., Hunter, E. & Teich, N. in *RNA Tumor Viruses* (eds Weiss, R., Teich, N., Varmas, H. & Coffin, J.) 513-648 (Cold Spring Harbor Laboratory, New York, (1982).
8. Miyata, T. *et al.* *Population Genetics and Molecular Evolution* (eds Ohta, T. & Aoki, K. (Japan Scientific Societies, Tokyo and Springer, Berlin, (in the press).
9. Toh, H., Hayashida, H. & Miyata, T. *Nature* **305**, 827-829 (1983).
10. Chiu, I.-M., Callahan, R., Tronick, S. R., Schlom, J. & Aaronson, S. A. *Science* **223**, 364-370 (1984).
11. Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. *Nature* **293**, 543-548 (1981).
12. Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3618-3622 (1983).
13. Majors, J. E. & Varmas, H. E. *J. Virol.* **47**, 495-504 (1983).
14. Donehower, L. A., Huang, A. L. & Hager, G. L. *J. Virol.* **37**, 226-238 (1981).

Prediction of protein structure from sequence

SIR—In a recent News and Views article, Stephen Harrison¹ urges caution "in predicting aspects of protein structure from sequence, what we need to recognize are entire domains or significant sub-domains—not just elements of secondary structure". He also points out "The poor correlation between local sequence and secondary structure has been aptly demonstrated by Kabasch and Sander², who show that the same pentapeptide in different proteins has fundamentally different backbone configurations in each".

I wish to point out two items which should caution the reader against accepting such advice without question.

In our original paper on the prediction of protein secondary structure³ (also see ref. 4), it was pointed out with great care that any specific sequence, for example five residues, will assume a conformation dependent upon the neighbouring sequences on both the amino and carboxyl termini until a sequence is reached, of usually four residues, which will either change conformation (β -turn), or will be unable to maintain a secondary structure [$\langle P_{\alpha} \rangle$ or $\langle P_{\beta} \rangle < 1.0$]. Thus it is an expected observation, rather than a novel finding, that a pentapeptide, joined to different sequences may assume dissimilar conformations. Thus the Kabasch and Sander² discovery, used to discredit predictive schemes, reaffirms rather than discredits the predictive algorithm. Harrison's¹ emphasis on domain structure rather than secondary structure to evaluate biological functional units has been expounded with great elegance by Rossman⁵ and is a well accepted theme. However, domains are constructed by in-

teractions of secondary structures, so let us not put the cart before the horse. To understand the formation of domains one must know the secondary structures first. Thus before one has a complete three-dimensional structure of a protein, derived from X-ray crystallographic studies, it can be extremely useful to predict secondary structures and look for homologies with other proteins to find interesting relationships.

I agree with Harrison¹ that the paper of Kabasch and Sander² should be carefully read as it displays how a false conclusion can be drawn from an interesting fact.

GERALD D. FASMAN

Graduate Department
of Biochemistry,
Brandeis University,
Waltham,
Massachusetts 02254, USA

1. Harrison, S. *Nature* **313**, 736-737 (1985).
2. Kabasch, W. & Sander, C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1075-1078 (1984).
3. Chou, P. & Fasman, G. D. *Biochemistry* **13**, 222-245 (1974).
4. Chou, P. & Fasman, G. D. *Adv. Enzym.* **47**, 45-148 (1978).
5. Rao, S. T. & Rossman, M. G. *J. molec. Biol.* **76**, 241-256 (1973).

On the possibility of inductive probability

SIR—We here make a second attempt to explain the reason for our hesitation in accepting the Popper-Miller thesis¹. We agree with all of their mathematical relations. Our problem is with the interpretation. We offer again two reasons and frame them here without using the negative signs of our first letter², which appear to have drawn excessive attention to themselves.

First, consider the (correct) result for $p(e) < 1$

$$s(h \leftarrow e, e) \equiv p(h \leftarrow e|e) - p(h \leftarrow e) < 0 \quad (1)$$

Popper and Miller identify the ampliative (or "loosely speaking inductive") part relative to any evidence e of any hypothesis h , with $h \leftarrow e$ given e . They infer from (1) that that part of h which goes beyond e is always countersupported by e . (We use their wording in both these sentences.) Our difficulty with this is that $s(k \leftarrow e, e) < 0$ for all values of k and in particular for $k = h$ and $k = \bar{h}$. One might have expected that if the inductive part relative to e of h is countersupported by e , then it would at least be possible for the inductive part relative to e of $\bar{h}$ to be supported by e . But on Popper and Miller's interpretation the inductive part relative to e of both h and $\bar{h}$ is always countersupported by e . This suggests to us that there is something astray with this interpretation, and it led us to observe in ref. 2 that a different definition of terms or some restriction on their interpretation seems to be required. Our interpretation of (1) was explained in our comment on our theorem A, and is rather simple: Since $p(h \leftarrow e)$ still allows $\bar{e}$ to occur, (1) shows that this probability is reduced if $\bar{e}$ is ruled out.

Second, one can understand the import-

ance of discussions of the extent to which 10 or 20 supportive occurrences e support a hypothesis h as a general statement, as is done in statistical theory. If a penny comes up "heads" 10 times ($e_1, e_2, \dots$) is it a penny with two heads (h)? There is a structured relation between e and h . It is, however, difficult to see how statements about induction can go very far if they are based only on probability algebra which leaves h and e as arbitrary propositions. In our letter we noted that the arguments in this correspondence had used only probability algebra, and that they involve the handicap of being independent of any inferential relationship between e and h .

Mathematically speaking, we would say that the support function $s(x, y)$ is indeed useful, as we noted in our comments on our theorem B. It is its division into the suggested "deductive" and "nondeductive" or (loosely speaking) inductive components (and particularly, the interpretation of the letter) that has given rise to our difficulties.

We are grateful to Karl Popper and David Miller for the courtesy of some private exchange of opinions.

J. WISE

P. T. LANDSBERG

University of Southampton,
Southampton SO9 5NH, UK

1. Popper, K. & Miller, D. *Nature* **302**, 687-688 (1983); **310**, 434 (1984); **315**, 461 (1985).
2. Wise, J. & Landsberg, P. T. *Nature* **315**, 461 (1985).

The form of maps in the brain

SIR—Further to Professor Ettlinger's pertinent comments on flying fox somatotopy¹, may I suggest that, if we accept the approach of metric tensor modelling for cortex as well as cerebellum², there would seem to be no requirement for any particular coordinate reference frame, provided that the sensorimotor connections to the central nervous system always preserve the tensorial relationships. Admittedly, my personal interpretation of the metric tensor model leaves me asking why we (and presumably some other vertebrates) need to construct a cartesian four-dimensional space-time at all.

H. STOWELL

ERBP Laboratory,
120 Nature Creek SW,
Milledgeville,
Georgia 31061, USA

1. Ettlinger, G. *Nature* **315**, 285 (1985).
2. Pellionisz, A. & Llinás, R. *Neuroscience* (1977-1985).

Selfish DNA and the origin of introns—Correction

IN this piece of Scientific Correspondence (by T. Cavalier-Smith, *Nature* **315**, 283-284; 1985) the word 'not' was omitted from line 14 on page 284. The sentence should begin: 'But nuclear tRNA does not depend on specific intron sequences ...'

Approaching the past

Colin Renfrew

Archaeological Explanation: The Scientific Method in Archeology.

By Patty Jo Watson, Steven A. LeBlanc and Charles L. Redman.

Columbia University Press: 1984. Pp. 309. \$26.50 (United States), \$34.50 (elsewhere).

Anthropological Archaeology.

By Guy Gibbon.

Columbia University Press: 1984. Pp. 455. \$29.50 (United States), \$38.50 (elsewhere).

ONE of the principal objectives of the "New Archaeology" of the late 1960s was to be "scientific". This meant less the use of various technical resources drawn from the natural sciences, such as radiocarbon dating or trace element analysis, than the attempt to place the reasoning behind the subject on a systematic basis.

The most persuasive of the New Archaeologists was Lewis R. Binford. Together with the first generation of his students, he sought to enlarge the scope of archaeology by asking new questions of the archaeological record, by developing techniques (such as efficient sampling strategies) which could answer specific questions once these had been clearly formulated, and by attempting to find more satisfying explanations for past events. These explanations were framed in terms of culture process: that is to say, in terms of general perceptions of the way various aspects of human societies have tended to operate. Binford and his colleagues were critical of many of the more traditional historical explanations, which were seen as frequently particularistic, tending merely to illuminate the case in question with a wealth of circumstantial detail, and often explaining the major events of the past in terms simply of the intentions of the principal actors, the Great Men of history.

The first generation of New Archaeologists emphasized the need to make explicit the bases of archaeological inference, to avoid judgements based on purely subjective assessment, and to find ways of testing hypotheses about the past, wherever possible by the use of fresh data. All of this added up to what the late David Clarke rightly termed the "loss of innocence", implying the development of a critical self-awareness of the nature of the discipline. The subject has continued to develop steadily since those early days, and has indeed met some of the hopes of those advocating an "explicitly scientific approach".

This phrase, *An Explicitly Scientific Approach*, formed the subtitle of the fore-runner of the first of the books under review, which was originally entitled *Explanation in Archeology* and published in 1971. It was an important book, seeking in a systematic way to apply the insights offered by the philosophy of science to aid archaeologists in making their subject

more scientific. This should, one might think, have been a royal road to success, but it proved instead something of a detour over philosophically shaky terrain towards rather sterile ground. The authors were open to criticism (as some reviewers indicated at the time) for following rather too readily the view propounded by the philosopher of science Carl Hempel. Hempel argued that nearly all explanation, whether in history or in the sciences, rests ultimately on the formulation of universal laws. His Deductive-Nomological (D-N) model of explanation offered an ideal to which all sciences (even historical ones) were thought to aspire.

Rather than follow this portentous mentor, the authors would have done better to ask what, in various good explanations known to them in the field of archaeology, is in fact the underlying structure of the explanation, rather than what it *ought* to be. What was clear to several archaeologists in 1971—that Hempel was an unsatisfactory guide—is clear to many philosophers of science today. Within that discipline what was once a widely accepted commonplace, the D-N model of explanation, is now seen as naive and problematic, and of much narrower application.

The present work, published 13 years later, has certainly learnt something from its initial *mésalliance* with a dogmatic philosophy of science. Different forms of Cover Law model are discussed in the first chapter, entitled "The Logic of Science, the Nature of Explanation and Archeology as Science". And happily the book moves on from these matters to talk about various aspects of archaeology in an informed and stimulating manner. Systems theory and ecology are discussed in succeeding chapters, and then the archaeological record and the design of archaeological research, and archaeological analysis, with a final chapter entitled "Archeology and Society".

Unfortunately, however, the authors have not been able to escape from the influence of Hempelian thought, and the tone of the book is firmly established in the first chapter. I find their view of science, as presented there, insensitive to the problems associated with explaining the behaviour of highly complex systems, as most living systems are. The D-N framework and statistical laws may work

well for physics, or for fields such as thermodynamics or psephology where very high numbers of unitary individuals make a probabilistic approach appropriate. But I doubt whether a molecular biologist or a meteorologist would recognize his or her science in these formulations. In fact the formal analysis of very complex systems has developed considerably in recent years, and we can begin to see that, even in the hard sciences, the trajectory of the system (that is to say its past history) as well as its present state may be relevant to its future behaviour. Fields such as catastrophe theory and the study of non-equilibrium systems are influencing archaeology as well as other sciences. The criticism to be made is therefore not necessarily the familiar one that human affairs cannot be treated scientifically, but rather that the vision of science offered here is not well developed for the explanation of the behaviour of highly complex systems.

In general, I think that the philosophy of science has served archaeology very badly over the past 10 to 15 years. The aim of being scientific, of setting explanations within an explicit framework of coherent theory, of developing a valid epistemology, is incontestably worthy. But so far most of the contributions from philosophers have been of a Procrustean inflexibility, seeking to impose on the subject under study, in this case archaeology, a series of preformulated precepts which have later themselves been shown to be of limited value. It now seems to me that to be a good philosopher of archaeology, and to produce insights into the nature of explanation in archaeology, it may pay to be a good archaeologist first.

This book, alas, continues to suffer from some of the defects of its predecessor: it tells us first how archaeology would look if it were scientific, and then tries to make the image fit. This is not only poor archaeology: it is poor philosophy of science. Perhaps it is even poor philosophy. I find it disconcerting that the first chapter, which attempts to define the nature of explanation in archaeology, and does so in terms of universal laws, offers very few examples of this kind of explanation in archaeology, and none at all which relates to human behaviour.

The book goes on to discuss several aspects of archaeology in an interesting way. It is invariably well-informed, and certainly worth reading. I share its general objectives: archaeology did need to be set on a better basis, and has benefited from the serious discussions about its nature over the past 15 years. But I am doubtful of the general validity of the view of science which is offered here.

Anthropological Archaeology is a new work which, likewise, sets out to tell us what science is, and what the philosophy of science is, before telling us very much about archaeology. It is serious, and

ANNOUNCING NEW PUBLICATIONS

**Continuing the
Nuclear Dialogue
Selected Essays by
ALVIN M. WEINBERG
ANS order #690009
\$25.00 sb**

and

**Nuclear Criticality Safety
by RONALD A. KNIEF
ANS order #300020
\$37.50 sb**

**order today from the
AMERICAN NUCLEAR SOCIETY
555 N. Kensington Ave.
La Grange Park, IL 60525
(312) 352-6611**

Reader Service No.40

IBM PC OWNERS*

**A DIGITIZER WITH INTEGRATED
MEASUREMENT AND GRAPHICS
FOR ONLY \$495**

**A versatile digitizing pen; accurate,
reliable and rugged.**

A Scientific Measurement System.

Linear • perimeter • area • X, Y
coordinates • angles • statistics:
standard deviations, correlations,
regression lines, transforms.

A Pocket Drafting System. Multiple

CAD-like aids: auto lines, boxes,
ellipses, multi-point curve fitting,
enlarge, shrink, invert • color lift-off
• symbol libraries • many type fonts.

Quick Lecture-Slides. Bar-graphs,
line graphs • with auto-edit • sizing
• conversion • labeling • coloring
• programmed for Polaroid Palette.

Display and Printout on dot matrix or
color printers • choice of size, shape.

The Digital Paintbrush System from
Jandel Scientific carries a 15-day
money-back guarantee. Call toll-free
1-800-874-1888 or in California
415-331-3022.

Jandel Scientific

2656 Bridgeway, Sausalito, CA
94965.

*Apple version \$299

Reader Service No.41

rather long. But although richer in examples (and illustrations) than *Archeological Explanation*, to my mind the author does not always understand his own examples very well, giving abbreviated summaries of important work which do not always catch the essence of the original. For instance I am not satisfied that the outline of the "wave of advance" model for the spread of agriculture (pp.262-265) shows how a delightfully simple dynamic formulation is derived from a very plausible set of explicit assumptions and initial conditions.

There are also some confusions, which are reflected in the glossary. "Diachronic analysis" — which implies analysis involving the passage of time and contrasts with synchronic analysis — is defined as "in the systems approach, analysis that is concerned with the mechanisms rather than the structural end products of changing subsystem relationships". This seems needlessly complicated. Survivorship curve is defined as "estimates of the percentage of individuals in subgroups of a theoretical population of 100 people that will still be alive at the end of a five year period". (Query: why specify five years?) "Tool kits" are said to be "spatial associations of archaeological materials that are assumed as a working hypothesis, to be meaningfully related to a specific task". That lacks precision, as well as clarity, and could apply as well to Stonehenge or the Pyramids of Egypt as to the subject intended.

Anthropological Archaeology does cover a number of issues seriously and thoughtfully. Its intentions, like those of *Archeological Explanation*, are thoroughly honourable. Many processual archaeologists, myself included, would share the author's aim of treating the subject in a thorough and systematic way. But this book does not convey the clarity of thought which anthropological archaeology should employ.

These two works reflect the real concern of the contemporary archaeologist to come to terms with the epistemological basis of the subject, and to develop coherent theoretical approaches. For this they will command respect. But neither they, nor the philosophers of science on whom they rely, adequately convey the lesson that different kinds of systems behave in different ways, and that complex systems can only rarely be described, and their behaviour explained, in terms of universal laws which go much beyond the obvious, the banal or the very general. The moral seems to be that the philosophy of archaeology, or of any specific science, is too important a subject to be left to the dictates of the generalist philosopher of science. □

Colin Renfrew is Disney Professor of Archaeology at the University of Cambridge, Downing Street, Cambridge CB2 3DZ, UK, and a Fellow of St John's College, Cambridge.

The shape of things to come

John D. Barrow

Mathematics and Optimal Form. By Stefan Hildebrandt and Anthony Tromba. W.H. Freeman: 1985. Pp.215. £20.95. To be available in the United States outside the Scientific American Library later this year.

WHAT do dirty washing and the design of the Great Pyramid have in common? The sort of tenuous "connection" that might generate an entire BBC television series perhaps? But there really is a connection between soap-suds and skeletons, honeycombs and Hero, the Queen of Carthage and the Olympic stadium in Munich. All are linked by the intriguing subject of optimal mathematical form, which was first discussed as a controlling factor in the natural world by D'Arcy Thompson. Here, two mathematicians who have made important contributions to the calculus of variations and minimal surfaces have produced a well-written and profusely illustrated study of the influence of extremum principles in nature.

This subject has a curious pedigree. The first modern variational principle, the so-called "Least Action Principle", was framed by Moreau de Maupertuis in 1744, although "minimum" principles had been employed by Hero to demonstrate the law of optical reflection nearly 1,700 years earlier. Maupertuis was influenced by Leibniz's doctrine of the "best of all possible worlds" and sought to overcome the defect that it appeared to be a statement of comparative reference about a unique entity. How can you say the observed laws of motion are optimal when you know of no others? Maupertuis pointed out that a certain combination of mass, velocity and the distance traversed was always minimized by the actual behaviour of a dynamical system, a requirement equivalent to imposing Newton's laws of motion. Maupertuis claimed that this displayed precisely in what sense the laws governing our world were optimal, and the inferior alternative worlds were those whose laws of motion gave paths of non-minimal action. This provided the basis for his teleological interpretation of nature. He even regarded the Least Action Principle as the most powerful proof for the existence of God. (A hundred years later a few theologians tried to maintain the doctrine of special creation in the face of fossil finds by claiming that the "failed" species revealed there were the products of the paths of non-minimal action!)

Hildebrandt and Tromba include much of the history of the calculus of variations and the appearance of optimization problems in the ancient world, for example the challenge to Dido of Carthage that she

could colonize only the area of land enclosed by the skin of an ox. Having split the skin into the thinnest strips possible, what is the optimal shape to enclose? The core of the book explains aspects of mathematical research into minimal surfaces and a variety of optimization problems involving processes occurring on unusual geometries and topologies. The results are illustrated by examples from biology, architecture, foams and bubbles, fluid flow and crystallography. Even professional mathematicians will find much that is new to them in these chapters — besides providing grist for popular lectures, they might provoke inclusion of options on the rigorous theory that underlies the subject in undergraduate mathematics courses.

Physicists will detect several absent topics that deserved inclusion in a book of this type. For example the discussion of geodesics in general relativity is too brief to be useful and misses the key contrast between Euclidean space and Minkowski's space-time: in Minkowski space-time there exists a longest world-line joining any two points, but no shortest one. The opposite is true in Euclidean space.

Altogether this is a most enjoyable book — a beautiful illustration of the interplay between mathematics and the natural world, interestingly and carefully written. □

John D. Barrow is a Lecturer in Astronomy at the University of Sussex, Brighton BN1 9QH, UK.

Spirits of Victorian England

Ivor Grattan-Guinness

The Other World: Spiritualism and Psychical Research in England, 1850–1914. By Janet Oppenheim. Cambridge University Press: 1985. Pp.503. £25, \$44.50.

ENGLISH Victorian thought continues to fascinate historians, not only for the quantity and quality of the work achieved but also for the range and complexity of philosophical positions attendant upon it. Not for them the quick lazy positivism of our modern science: this was natural philosophy, the study of nature in all her physical manifestations, and with the ontological and religious aspects set in place. One feature of these endeavours was the prominence accorded to study of the mind

and its means of communication, both in orthodox contexts and in the realms of "spiritual" activity. "The word *spiritual* is one of the most difficult and important terms in nineteenth-century thought", wrote F.M. Turner in his *Between Science and Religion: The Reaction to Scientific Naturalism in late Victorian England* (1974), and various historians have recently developed this theme in the context of psychical research, especially B. Inglis in *Natural and Supernatural* (1977). Now we have the best study to date.

Janet Oppenheim, a young American scholar about whom the publisher tells us nothing, has ransacked the publications of the period, a wide range of manuscript collections and the relevant secondary literature to produce not only another account of the rise of psychical research in the late nineteenth century but also a detailed statement on the varieties and place of spiritualism during that period. Her book falls into three parts. "The Setting" describes in two short chapters the activi-

ties of various mediums and the wide spread of spiritualist publications and societies in England at that time. "A Surrogate Faith" explains with great lucidity the plurality of positions embraced by spiritualism — the survival of spirit after death was the common assumption, but a wide range of views was adopted by different groups — and then the complicated relations between spiritualists, the principal workers within the Society for Psychical Research (SPR), which was founded in 1882, and theosophists. After a welcome suite of portraits and illustrations, there then comes the excellent third part, unhappily entitled "A Pseudoscience", in which the places of psychical research and spiritualism are described for three areas of science: in turn, psychology and phrenology (the latter a special *Scottish* interest, as the author might have pondered); evolution and medicine; and physics.

Four hundred pages of beautifully composed text are followed by 90 pages of end-notes, full of interesting information and references; however, they lack proper cross-references, so that it can be hard to find the full details of an item.

The richness and complications of Victorian thought cannot be fully conveyed in a review, although the figure (see left) and its attendant commentary include some further explanation. I focus, therefore, on the last part of the book, which is of most interest to scientists today. Sensibly, Oppenheim chooses to present the variety of positions by describing *seriatim* the views of major figures concerning spiritualism, religion, psychical research and their specialist areas of knowledge: wholehearted spiritualists such as Alfred Russel Wallace and William Crookes; men of a more cautious or vacillatory stance, such as G. J. Romanes, J. J. Thomson and Oliver Lodge; and those typified by Henry Sidgwick and his wife Eleanor, who could not subscribe to any established religious position (and in their cases also became opposed to spiritualism). The only group omitted from the survey are logicians, which is a pity, since at that time logic and psychology were closely linked. Augustus de Morgan and his wife are mentioned, but they lack the company of Mrs Boole, propounder of her late husband's algebra and also founder council member of the SPR; Lewis Carroll and John Venn; and W. E. Johnson, brother and mathematical assistant to the first research officer of the SPR, Alice Johnson.

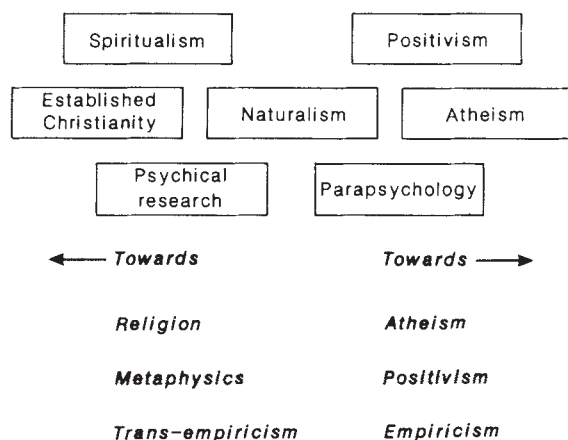
The intertwining of the study of nature, the belief in survival and the pursuit of the paranormal that characterized a large part of Victorian thought is re-created in this book to an extent not previously achieved by an historian. Oppenheim has produced an altogether superior study. □

Ivor Grattan-Guinness is a council member of the Society for Psychical Research, 1 Adam & Eve Mews, London W8 6UG, UK.

REPRESENTATION of the spectrum of philosophical positions adopted by Victorians on religious and scientific questions. The boxes are placed on separate lines merely for convenience: the linear form adopted is a considerable over-simplification (so that, for

example, theosophy cannot fairly be indicated on it), but it conveys some impression of the distinctions and overlaps. "Naturalism" was a term used at the time to describe views which usually preferred Nature to God, facts to belief, reason to faith, and

matter to spirit. To one side lies "psychical research", referring to views normally held by leaders of the SPR (which, however, held no *corporate* opinion); to the other is "parapsychology", used in the narrow sense to refer to the experimental, laboratory-orientated studies associated from the 1930s on with J. B. Rhine and his followers. The linear diagram, then, also has a chronological overtone, from left to right. For further details see *Research in Parapsychology* 1982 pp. 283–304 (Scarecrow Press, Metuchen, NJ, 1983).



Elements of disaster

John B. Whittow

Volcanic Hazards: A Sourcebook on the Effects of Eruptions. By R.J. Blong. Academic: 1984. Pp.424. \$66, £49.50.

IN A period when such environmental disasters as drought in Ethiopia and flood in Bangladesh make headline news, it has now been generally acknowledged that many Third World communities have become so "marginalized" that the slightest aberration in their local atmospheric or geophysical systems may trigger a disaster. Records suggest that there has been no increase in either the magnitude or the periodicity of hazardous events, but death tolls continue to rise.

Bare statistics disguise a significant trend, however, for while the loss of life from natural disasters has increased in the developing world, deaths in the developed world have fallen. Most observers claim that the dichotomy results from the "technological cushion" enjoyed by developed nations, where scientific knowledge, organized planning and advanced warning systems have combined to alleviate the threat and ease the impact of natural hazards. Critics highlight the distinction between those societies which can choose from a long list of structural and non-structural adjustments and those which have no choice. The latter have limited expertise and local infrastructure, little or no opportunity to insure against disaster, and even if warnings could be disseminated the lack of mobility inhibits evacuation: they must bear the loss of life and property.

Any new publication dealing with natural hazards must be viewed in this light. Increasingly, the value of such books is judged by the degree to which they offer to help reduce the vulnerability of threatened populations, whatever their economic or social status; it is no longer enough simply to catalogue and explain the physical phenomena of natural hazards. One is encouraged, therefore, to discover that *Volcanic Hazards* has this humanitarian concern as one of its three aims. Its other objectives, of chronicling and analysing the effects of the hazard, follow the traditional approaches adopted by the plethora of books on the subject which have appeared over the past decade.

Volcanic hazard research has progressed far beyond such pioneering works as Symons and Lacroix who, respectively, described the famous eruptions of Krakatau and Mont Pelée almost a century ago. Even the modern global syntheses of Bullard¹ and Francis² have rapidly

been overtaken as earth scientists have become more concerned with effects than with causes. Following the valuable work of Walker³ on volcanic prediction, and of Sheets and Grayson⁴, who examined the relationship between volcanoes and human ecology, there has been an urgent need for a single volume that would bring together the mass of literature on volcanic eruptions in a clear and concise form. R.J. Blong has provided such a volume. Here is a well-written "sourcebook" (the author's own subtitle) which not only examines all of the well-chronicled volcanic events from Santorini (1500 BC) to Mount St Helens (AD 1980), but also provides, at the end of each chapter, an appraisal of the impact of volcanic activity on various aspects of the environment.

In Blong's assessment of deaths and injuries caused by volcanic eruptions, several points emerge. For instance,

activity have resulted from starvation and disease. It is a pity, therefore, that while the devastating effects of tephra falls, pyroclastic flows etc. on crops and livestock are critically discussed at local scales, Blong says virtually nothing about the influence of volcanic dust veils on world climates and thus on harvests. Unfortunately, he believes the topic to be beyond the scope of his book.

There are no such omissions in the sections dealing with the impact of volcanic eruptions on buildings, public utilities, communications and economic activity. When it comes to property losses we discover, not surprisingly, that it is the developed world that has borne the brunt. Almost half of the book is devoted to an analysis of the effects on the built environment, and although worldwide case studies are included Blong relies heavily on the Mount St Helens disaster.

This is a welcome synthesis of information on the 1980 eruption, and one which puts the economic effects in their true perspective. We learn, for example, that local employment actually increased and that the economic impact was less than that of the Boeing recession. One suspects that were it not for the mass of volcanic perception studies generated by Mount St Helens, the valuable chapter on social aspects of eruptions would never have appeared. Like other hazard perception studies, it illustrates the impossibility of producing a regional disaster plan without knowing something of the likely behavioural responses of decision makers as

well as of those at risk. Without such knowledge disaster relief funding, evacuation plans and survival rates are difficult to assess.

Because it offers broad outlines for risk assessment, *Volcanic Hazards* will be invaluable to administrators, engineers and planners in developing and developed societies, as well as appealing to all earth scientists concerned with volcano monitoring and hazard delineation. A careful balance has been maintained in the treatment of man-environment relationships, the examples are well-researched, clearly explained and superbly referenced, and the whole is supported by many high-quality photographs, maps, diagrams and tables. This is an elegant volume which satisfies a longstanding need in the literature of natural hazards.

1. Bullard, F. M. *Volcanoes of the Earth* (University of Texas Press, Austin, 1976).
2. Francis, P. *Volcanoes* (Penguin, Harmondsworth, 1976).
3. Walker, G. P. L. *Geol. Soc. Lond. Misc. Publ.* 3, 23-41 (1974).
4. Sheets, P. D. & Grayson, D. K. (eds) *Volcanic Activity and Human Ecology* (Academic, New York, 1979).

John B. Whittow is Senior Lecturer in the Department of Geography, University of Reading, Reading RG6 2AB, UK.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Destruction by volcano — *The Last Days of Pompeii*, by Brüllow.

deaths due to volcanic activity are contrasted with those resulting from other natural hazards and it is shown, for example, that the victims of earthquakes outnumber volcanic fatalities by a ratio of 12:1. Of even greater note are the figures indicating that Indonesia suffers no less than 67.3 per cent of the world's volcanic casualties, compared to a combined total for New Zealand and the Mediterranean countries of 2 per cent, despite the fact that all three regions are together regarded as the most volcanically active in the world. The statistics can, of course, be explained in part by the respective population densities of the three areas. But when a developed country such as Japan (just as over-crowded as Indonesia) is taken into account, one has to conclude that Japan's 8 per cent of the total casualties must be explained by its "technological cushion" and its reduced vulnerability to hazard-induced crop losses and subsequent famine. Indeed, in the context of food shortages in the Third World, the chapter dealing with volcanic effects on agriculture is of great importance, in particular because, as Blong shows, since 1600 some 40 per cent of deaths from volcanic

Superfluidity in neutron stars

David Pines & M. Ali Alpar*

Department of Physics, University of Illinois at Urbana-Champaign, 1110 W. Green, Urbana, Illinois 61801, USA

Recent developments have radically changed our understanding of the dynamics of neutron-star superfluids. We review astrophysical evidence for the existence of at least three distinct superfluids inside neutron stars from the observed relaxation of the rotation and spin-down rates of the pulsars Vela, Crab and PSR0525+21, following sudden jumps (glitches) in these quantities. We also discuss the effects of the superfluid interior in the accreting neutron-star X-ray sources.

SUPERFLUIDITY in neutron stars was first suggested^{1,2} before the observation of such stars; the suggestions were as remarkably prescient as the first theoretical proposals of the existence of neutron stars and pulsars³⁻⁵. The discovery of radio pulsars and accreting X-ray sources placed neutron stars firmly in the realm of observational astrophysics. Subsequent accumulation of timing data revealed that radio pulsars experience sudden period changes ('glitches'), and their relaxation following such a glitch is too slow to be explained in terms of the viscous processes of normal matter. These timing observations provided the evidence for the presence of cosmic superfluidity and motivated the theoretical work on the rotational properties of this superfluid⁶. As is well known from laboratory experiments, the rotational state of a superfluid is determined by a distribution of quantized vortex lines, just as in a hard superconductor the magnetic field configuration is determined by a distribution of quantized flux lines. The motion of these vortex (or flux) lines is governed by their interactions with normal matter, including the possible pinning of the vortex lines to preferred sites in an inhomogeneous medium, a phenomenon well known in studies of hard superconductivity. The subject of this review is the fascinating extension of these basic concepts of superfluidity to the largest superfluid systems encountered in nature, neutron stars. Several physical effects peculiar to the dense state of matter in neutron stars play a key role in this extension.

The salient features of the current picture are that: (1) the core superfluid, which comprises the bulk of the moment of inertia, rotates rigidly with the crust of the star despite being in the superfluid state; (2) because of vortex pinning the crust superfluid, which contributes $\sim 10^{-2}$ of the star's total moment of inertia, can be out of equilibrium with the observed crust for timescales of weeks, months and years, and is responsible for a variety of interesting timing behaviour, primarily pulsar glitches and post-glitch relaxation; and (3) the observed post-glitch relaxation times provide a means of estimating the interior temperature of pulsars.

Structure of neutron stars

The details of neutron-star structure are quite sensitive to the still imperfectly understood neutron matter equation of state at densities greater than $\rho_0 \approx 2.8 \times 10^{14} \text{ g cm}^{-3}$, the density of nuclear matter⁷. If the basic hadron equation of state is comparatively stiff (see ref. 8 for a recent review), the cross-section of such a star would be that shown in Fig. 1. Beneath an atmosphere only a few metres thick one finds an outer crust, 0.5-km thick, of increasing density ($7 \times 10^6 \text{ g cm}^{-3} \leq \rho \leq 4 \times 10^{11} \text{ g cm}^{-3}$), containing a solid array of nuclei and a highly degenerate relativistic electron plasma⁹. The inner crust begins at $\rho \sim 4.3 \times 10^{11} \text{ g cm}^{-3}$ and extends a few km until a mass density $\rho \sim 2.4 \times 10^{14} \text{ g cm}^{-3}$

is reached. It contains a highly degenerate superfluid neutron liquid as well as the lattice of increasingly neutron-rich nuclei and relativistic electrons. The quantum liquid interior of the star is expected to consist mainly of superfluid neutrons, with an admixture of a few per cent of superconducting protons and normal electrons. The constituents of neutron stars at densities well in excess of ρ_0 are uncertain (see ref. 10 for a review of some of the possibilities). It is possible that at densities a few times ρ_0 one may find an alternating layer structure involving a pion condensate^{11,12}. Quark matter is not expected¹⁰ because the maximum central densities of neutron stars ($\leq 5\rho_0$) are rather less than those at which such a state would be favoured. Because of the very high thermal conductivity of degenerate matter, the temperature, T , of the star is quite uniform, and is typically $\sim 5 \times 10^8 \text{ K}$ for a young neutron star such as the Crab pulsar. This would seem at first sight to be high, until one realizes that, because of the extraordinarily high densities involved, typical crustal melting temperatures are $\sim 10^{10} \text{ K}$ (1 MeV), whereas degeneracy temperatures for the quantum liquid constituents are still higher (5–50 MeV) and energy gaps for the neutron and proton superfluids may represent an appreciable fraction of their degeneracy temperatures. The crustal material thus resembles terrestrial solids in the liquid helium range, whereas the quantum liquids resemble liquid He in the millidegree range.

Neutron superfluids

The effective interaction between two neutrons is a combination of a strong short-range repulsion and a long-range attraction. At least two distinct pairing states of neutron superfluids are expected: (1) in the inner part of the crust, where the interparticle spacing is greater than the range of the repulsive forces ($\sim 10^{-13} \text{ cm}$), 1S_0 pairing dominates¹³⁻¹⁵. (2) at the higher densities ($\geq \rho_0$) characteristic of the quantum liquid interior, the combination of short-range repulsion and an attractive tensor force leads to a 3P_2 paired superfluid^{13,16}. The transition temperatures are shown in Fig. 2 as a function of density in the star.

Because neutron stars rotate, neither neutron superfluid will be spatially uniform; both contain an array of vortex lines parallel to the rotation axis, of sufficient number that on a macroscopic scale the neutrons will seem to be rotating as a rigid body. As the circulation per vortex is $\kappa = h/2m_n$, the density of vortices per unit area is $2m_n\Omega/\pi\hbar$ to mimic rigid-body rotation (where $\hbar$ is Planck's constant, m_n is the neutron mass, and Ω is the rotation frequency). This vortex density is $\sim 2 \times 10^5 \text{ cm}^{-2}$ for the Crab pulsar ($\Omega \approx 191 \text{ s}^{-1}$).

In the stellar crust, the normal cores of the vortices where the condensate wavefunction decreases to zero either may be pinned to the lattice of nuclei present there or may thread the spaces between them. Because the nuclei also contain superfluid neutrons, whether or not pinning occurs is a matter of comparative superfluid energetics. It is favourable for a vortex line to

* Present address: TBTAk, Research Institute for Basic Sciences, Gebze-Kocaeli, Turkey.

pin to a nucleus if the energy cost per particle to create its normal core, $\sim \Delta^2/E_F$, where Δ is the energy gap for superfluid neutrons and E_F their Fermi energy, is thereby reduced⁶. More precisely, the pinning energy per nucleus is^{17,18}

$$E_p = \frac{3}{8} \left[\frac{\rho_{\text{out}} \Delta_{\text{out}}^2}{(E_F)_{\text{out}}} - \frac{\rho_{\text{in}} \Delta_{\text{in}}^2}{(E_F)_{\text{in}}} \right] V \quad (1)$$

Here 'out' and 'in' denote local values of Δ and E_F for superfluid neutrons outside and inside the nuclei, V is the overlap volume of the vortex core and the nucleus of radius R_N ; the core radius is the coherence length of the outside neutron superfluid,

$$\xi = 2E_F/\pi k_F \Delta \quad (2)$$

Calculations^{15,19} suggest that the pinning layers, in which it is energetically favourable for vortex lines to go through the nuclei, extend from densities $\sim 10^{13}$ to $\sim 2 \times 10^{14} \text{ g cm}^{-3}$. With a pinning energy E_p , the maximum available pinning force per unit length of vortex line is

$$f_p = \frac{E_p(\rho)}{\xi(\rho)b(\rho)} \quad (3)$$

where ρ is the average density characterizing the pinning layer and b is the spacing between successive pinning centres along the vortex line.

The pinning energy and coherence length are determined by the equilibrium nuclear configuration²⁰ and the superfluid energy gap, while the distance between pinning centres depends on the physical circumstances under which pinning takes place. Pinning will be maximal in the strong pinning layers, where E_p is greater than a typical lattice energy, $E_L \sim Z^2 e^2 R_N / b_z$ (b_z is the lattice spacing), so that the pinning force is strong enough to dislodge nuclei from their equilibrium sites in the lattice, and b is comparable with b_z . The transition to weak pinning layers occurs when $E_p \leq E_L$. In the weak pinning layers, the vortex core pins only to those sites through which it passes; the spacing between pinning sites, b , is $\sim (E_L/E_p)b_z$, and the effective pinning energy is reduced to $\sim E_p(E_p/E_L)$ (ref. 18). Finally, where the coherence length, $\xi \geq b_z$, a vortex core can 'pin' to more than one nucleus at a time. Under these circumstances, which are expected in the higher-density regions of the pinning layers, moving a vortex line requires still less change in the pinning energy; one is in a region of superweak pinning¹⁹, in which the effective spacing, b , between multiple pinning sites is far larger than $(E_L/E_p)b_z$.

The proton superfluid

The basic interaction between isolated protons resembles that for neutrons, whereas the density of the proton liquid in the stellar core is some two orders of magnitude smaller than that of the neutron liquid; hence the interior proton liquid will superconduct in a 1S_0 pairing state²¹.

Because the critical field, $H_{c1} \geq 10^{15} \text{ G}$, is enormous compared with the field strengths $\sim 10^{12} \text{ G}$ that are found in pulsars, one might expect that the proton superconductivity would have a dramatic effect on the magnetic properties of the neutron star, through expulsion of the flux from the superconducting regions. It does not, because the electrical conductivity, σ , of the normal state is so large that the characteristic time for flux expulsion from macroscopic regions is typically $\gg 10^6 \text{ yr}$ as Baym, Pethick and Pines²² have shown. As the star cools, superconductivity nucleates with the magnetic field present. The London penetration depth is greater than the proton coherence length, so that the flux penetrates the superfluid region in a vortex state similar to that found in laboratory type II superconductors for $B > H_{c1}$.

Core neutron communication

The principal energy source of a pulsar is the rotational energy of the neutrons. How do the core neutrons, which make up the bulk of the neutron star, follow changes in the rotational state of the crust? Consider first the charged particles. The electrons and protons in the interior must co-rotate by electromagnetic

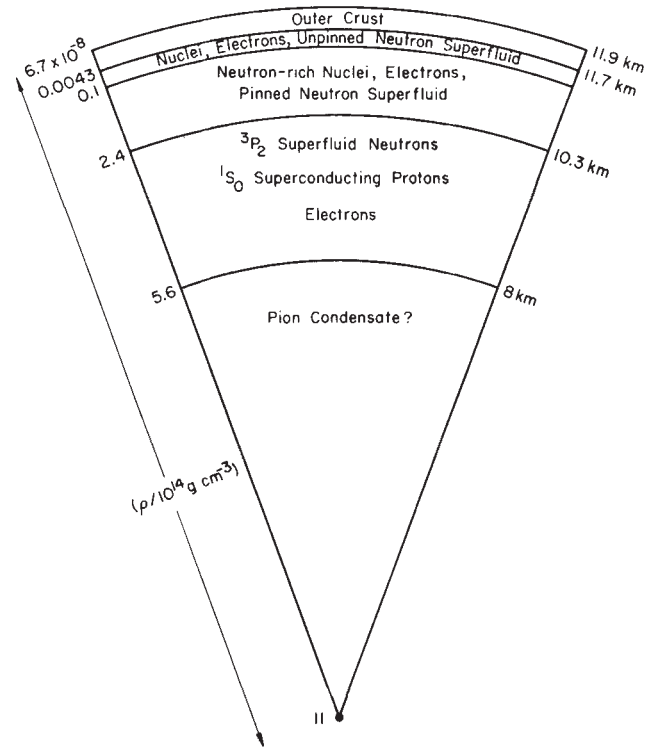


Fig. 1 Cross-section of a $1.4 M_{\odot}$ neutron star, calculated using the Bethe-Johnson equation of state.

coupling, irrespective of whether or not the protons are superfluid (any appreciable differential rotation would give rise to inordinately large magnetic fields). The charged particles in the crust and those in the interior communicate readily through the viscosity of the electrons and/or the magnetic field in the interior. Easson²³ finds that the core plasma responds to sudden changes in the crustal angular velocity with a magnetic spin-up time $\sim 10 \text{ s}$. The Ekman time for electrons is of the same order^{23,24}.

The rotational state of the neutron superfluid is determined by the distribution of its quantized vortices. Any change in the rotation rate of the crust is communicated to neutron vortex cores via electrons and protons. Thus, the dynamical coupling time of the neutron superfluid is determined by the timescale for the scattering of electrons against the normal cores of superfluid neutron vortices. For electron-neutron magnetic moment scattering, the charged particle-neutron coupling time is found to be $\sim 10^7 \text{ s}$. These considerations led to the conclusion that if both the neutrons and the protons in the stellar interior are superfluid, the relaxation time which characterizes the coupling between the crust and the interior superfluid neutrons is a macroscopic time, of the order of years²². This formed the basis for the proposal of the two-component model of crust-core coupling in neutron stars²⁵. More recent work^{26,27} has shown that there are mechanisms of intrinsic magnetization of the neutron vortices that result in a much stronger coupling of the neutron superfluid to the charged particles and to the crust.

Glitches and post-glitch relaxation

The inner part of the pulsar magnetosphere adjacent to the neutron star crust rotates with it. Thus, the observed pulse timing reflects the rotational behaviour of the crust. Sudden increases in rotation rate, termed glitches, have been observed from many pulsars. The Crab pulsar has had two prominent glitches²⁸ with $\Delta\Omega/\Omega \sim 10^{-8}$ – 10^{-9} . The old pulsar 0525+21 has exhibited a glitch of similar magnitude²⁹. The Vela pulsar has had six glitches^{30–32} with $\Delta\Omega/\Omega \sim 10^{-6}$ since 1969, and each of the pulsars 1641–45, 1325–43 and 2224+65 have shown one such giant glitch^{33–35}. The statistics of the Vela-type giant glitches is consistent with the hypothesis that all pulsars have them, although glitches become less frequent as pulsars evolve³⁶.

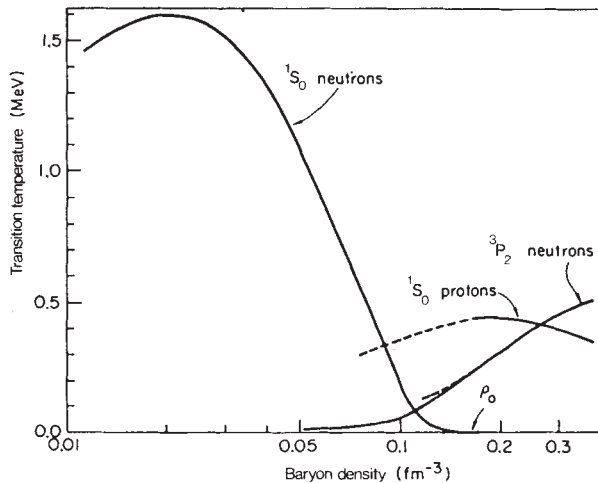


Fig. 2 Calculated superfluid transition temperature as a function of baryon density for neutrons and protons¹⁰. Dashed curves indicate extrapolations to lower densities.

Glitches are sudden in the sense that observationally their rise time seems to be less than a day. By contrast, the post-glitch relaxation in the timing behaviour of a pulsar, which has been observed in the Crab, Vela and PSR0525+21, has timescales from a few days to years.

Post-glitch relaxation is determined by the coupling of the crust to the superfluid neutrons. Let us write the torque acting on the crust of a neutron star rotating at angular velocity, Ω_c , as

$$I_c \dot{\Omega}_c = N_{\text{ext}} - N_{\text{int}} \quad (4)$$

I_c is the moment of inertia of the crustal lattice plus whatever other stellar material is rigidly coupled to it, N_{ext} is the external torque acting on the star and N_{int} describes the coupling of the crust to those superfluid neutrons that couple to it on observable timescales. In the two-component model of Baym *et al.*²⁵, the internal torque is ascribed to the core neutron superfluid of inertial moment I_s and angular velocity $\Omega(t)$, and is assumed to take the form

$$N_{\text{int}} \equiv I_s \dot{\Omega} = (I_c/I) (\Omega_c - \Omega) / \tau_c \quad (5)$$

characteristic of simple mutual friction between the crust and core (where I is the total moment of inertia and τ_c is a crust-core coupling time). After a sudden initial jump $(\Delta\Omega_c)_0$ in Ω_c , the post-glitch behaviour described by equations (4) and (5) is

$$\Omega_c(t) = \Omega_0(t) + (\Delta\Omega_c)_0 [Qe^{-t/\tau} + (1-Q)] \quad (6)$$

where $\Omega_0(t)$ is the extrapolated frequency in the absence of the glitch, $\tau = \tau_c(I_c/I)$ and $Q \approx I_s/I$. The two-component theory provided a good fit to the initial post-glitch data from the Crab and Vela pulsars. For both pulsars, however, the Q and τ values required to obtain a fit differed somewhat from one glitch to the next. Moreover, Boynton and Deeter³⁷ showed that one could not fit the noise in the timing data on the Crab pulsar using those Q and τ values which had provided a satisfactory explanation of the larger Crab glitches, while data accumulation and the detailed analysis by Downs³⁰ of the post-glitch behaviour of the Vela pulsar following its four glitches, and by Lohsen²⁸ and Demianski and Proszynski³⁸ of the Crab pulsar following its two major glitches, provided a wealth of new data for which the simple two-component model could not provide an adequate explanation.

The physical origin of the giant glitches in the Vela pulsar posed an even greater puzzle: starquake theory⁹, although plausible for the Crab pulsar glitches³⁹, could not explain the magnitude and frequency of the Vela glitches. The most promising alternative which emerged was the idea of a sudden catastrophic unpinning of the pinned vortex lines in the crustal neutron superfluid every few years as a result of pulsar spin-down^{40,41}. In the course of exploring the behaviour of the pinned superfluid

between giant glitches, it was concluded that not only could the pinned superfluid be responsible for the giant glitches, but that the entire dynamics of post-glitch behaviour was probably related to the pinned superfluid; in other words, the post-glitch behaviour was the response of the pinned crustal neutron superfluid to a glitch, and physical processes must exist which couple the core neutron superfluid rapidly (on timescales $\leq$ days) to the crust^{18,42}. Thus, a general theory of the coupling between the pinned superfluid and the crust, which occurs by vortex creep (the motion of the pinned vortex lines in response to changes in the crustal rotation frequency⁴³), was developed. A detailed re-examination of the coupling between the crust and the core neutron superfluid was carried out^{26,27} in which additional physical mechanisms which couple the core neutrons rapidly to the crust were discovered, while the vortex creep theory was found not only to provide an excellent fit^{19,44} to the timing data^{28-30,38}, but to offer the possibility of determining the internal temperature of pulsars.

Neutron star coupling

Sauls, Stein and Serene²⁶ considered the structure of the vortices in the 3P_2 pairing state in the neutron superfluid core. They found that because of the small spontaneous magnetization that appears in the region near the vortex core, the coupling is qualitatively different from the relaxation mechanism considered by Feibelman⁴⁵, who calculated the scattering of electrons by magnetic moments of neutrons in low-lying excited states in the core of the vortex line. This requires thermal excitations and yields a relaxation time which is $\sim \exp(\Delta^2/k_B T E_F)$ (where k_B is Boltzmann's constant), whereas the electron magnetic vortex relaxation time is comparatively insensitive to the gap and temperature, going as the inverse power of $\Delta(T)$; the latter process gives relaxation times of about a year.

Alpar, Langer and Sauls²⁷ have examined the change in the structure of a neutron vortex line that results from the interaction between the neutron and proton condensates in the quantum liquid interior. Superfluid drag induces a proton charge current around each neutron vortex giving rise to a magnetic field confined within a screening radius ~ 50 fm, but of magnitude $\sim 10^{15}$ G, far larger than the spontaneous magnetization discussed above. The resulting charge-vortex line relaxation time is also insensitive to temperature and is $\tau_v \approx 20 P(s)$, where P is the rotation period. The dynamical coupling time of the core superfluid is

$$\tau \equiv \frac{\tau_v}{x} \approx 400 P(s) \quad (7)$$

where $x \approx 0.05$ is the charged fraction of the neutron star core. This work provides crucial support for invoking only the crust superfluid in a model of pulsar glitches and post-glitch relaxation, which we now consider.

Vortex creep theory

The coupling between the pinned crustal superfluid and the crust is achieved by thermal creep of the vortex lines against pinning energy barriers, by analogy with flux creep in type II superconductors^{46,47}. In the steady state the pinned superfluid is coupled to the rest of the star and shares the effect of external torques. A glitch perturbs vortex creep away from its steady-state rate. Part of the pinned superfluid is thereby decoupled, and its gradual recoupling leads to observable behaviour. The sudden change in $\dot{\Omega}_c$ which accompanies the glitch is associated with sudden decoupling of the pinned superfluid region with an effective moment of inertia $\bar{I}_p$,

$$\bar{I}_p/I \equiv (\Delta\dot{\Omega}_c/\dot{\Omega}_c)_0 \quad (8)$$

and conservation of angular momentum gives

$$\bar{I}_p \delta\bar{\Omega}_p = I_c \Delta\Omega_c \quad (9)$$

where $\delta\bar{\Omega}_p$ is the mean decrease of the superfluid rotation rate and $\Delta\Omega_c$ is the magnitude of the glitch.

The angular velocity of the neutron superfluid is determined by the distribution of the pinned vortex lines. With azimuthal symmetry, the value of $\Omega(r) \equiv v(r)/r$ (where r is the cylindrical distance from the axis of rotation) is determined by the usual superfluid equations relating the macroscopic rotation rate to the distribution of quantized vortices:

$$\int_0^r (\nabla \times \mathbf{v}) \cdot d\mathbf{S} = \oint \mathbf{v} \cdot d\mathbf{l} = 2\pi\Omega(r)r^2$$

$$= \kappa \int_0^r 2\pi r' n(r') dr' \quad (10)$$

while vortex conservation leads to

$$\dot{\Omega} = -\kappa \frac{nv_r}{r} \quad (11)$$

Here, $n(r)$ is the number of vortices per unit area, $\mathbf{v}$ is the superfluid velocity, and v_r is the radial velocity of the vortex lines. From these equations it follows that $\Omega(r)$ can change only if vortices move. Their motion is described by an internal torque,

$$N_{\text{int}} \equiv \int_p dI_p \dot{\Omega}(r, t) \quad (12)$$

while the rate at which thermally activated vortex lines are able to tunnel between adjacent pinning configurations is

$$v_r = v_0 \exp \left[-\frac{E_p}{k_B T} \frac{(\omega_{\text{cr}} - \omega)}{\omega_{\text{cr}}} \right] \quad (13)$$

where T is the crustal temperature, $\omega \equiv \Omega - \Omega_c$ is the superfluid angular velocity relative to that of the crust, and

$$\omega_{\text{cr}} \equiv (\Omega - \Omega_c)_{\text{max}} = \frac{f_p}{\rho \kappa r} \equiv \frac{E_p(\rho)}{\rho \kappa r \xi b} \quad (14)$$

is the maximum superfluid lag which can be supported by the pinning force, defined in equation (3). The approach to steady state is characterized by a relaxation time, τ_i , proportional to the temperature⁴³,

$$\tau_i = \left(\frac{\omega_{\text{cr}}}{E_p} \right)_i \frac{k_B T}{|\dot{\Omega}_{\infty}|} \equiv \frac{\alpha_i T}{|\dot{\Omega}_{\infty}|} \quad (15)$$

where $\dot{\Omega}_{\infty}$ is the steady-state value of $\dot{\Omega}_c$ and $\dot{\Omega}_i$, and the subscript 'i' refers to the pinning layer in question.

Glitches are catastrophic unpinning events extraneous to the creep process which itself involves a continuous slow vortex motion. The internal torque produced by the neutron superfluid in response to a glitch, $(\Delta\Omega_c)_0$, takes the simple form

$$N_{\text{int}}(t) = \sum_i I_i \dot{\Omega}_{\infty} \frac{1}{1 + (e^{t_{oi}/\tau_i} - 1)e^{-t/\tau_i}} \quad (16)$$

where the I_i are the moments of inertia of the distinct pinning layers; the t_{oi} which appear in equation (16) are post-glitch delay times. For pinning layers where there is no glitch-associated vortex motion,

$$t_{oi} = \frac{(\Delta\Omega_c)_0}{|\dot{\Omega}_{\infty}|} \quad (17)$$

and in those regions where there is glitch-associated vortex motion,

$$t_{oi} = \frac{\delta\Omega_i}{|\dot{\Omega}_{\infty}|} = \frac{X_i \kappa}{2\pi r^2 |\dot{\Omega}_{\infty}|} \quad (18)$$

where $\delta\Omega_i$ is the glitch-associated change in the superfluid rotation frequency and X_i is the number of vortices that have moved through the region.

The vortex creep theory of post-glitch response differs from the two-component model²⁵: (1) It involves a number of distinct components which, however, make up only a comparatively small fraction ($\sim$ a few per cent) of the neutron superfluid, namely the pinned superfluid in the crust. This is borne out by the observed values of $(\Delta\dot{\Omega}_c/\dot{\Omega}_c)_0 \sim 10^{-3}$ – 10^{-2} (compare with

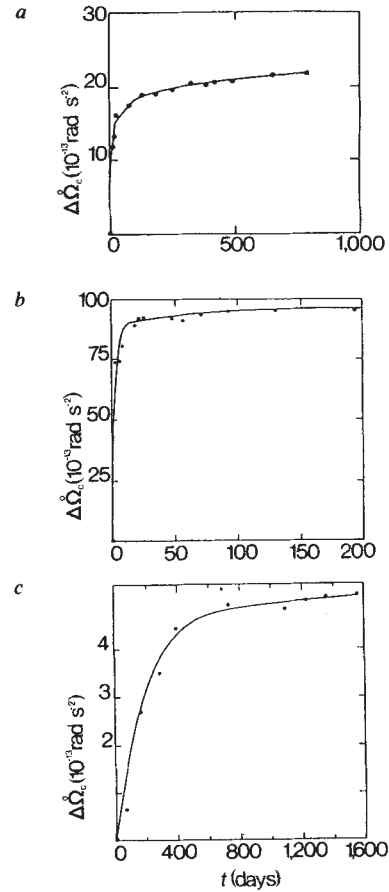


Fig. 3 Fits to the post-glitch $\dot{\Omega}_c(t)$. *a*, Representative post-glitch fit for the Vela pulsar¹⁹: $\dot{\Omega}_c(t)$ following the glitch of July 1978. *b*, Fit to $\dot{\Omega}_c(t)$ following 1975 Crab pulsar glitch⁴⁴. *c*, Fit to $\dot{\Omega}_c(t)$ following the glitch observed in PSR0525+21 (ref. 44). In each case, $t=0$ denotes the epoch of the first post-glitch timing observation and $\Delta\dot{\Omega}_c(t) \equiv \dot{\Omega}_c(t) - \dot{\Omega}_c(0)$.

equation (8)). (2) It is a fundamentally nonlinear theory of superfluid response; glitch-associated vortex creep depends exponentially on the size of the glitch (in regions in which no vortex motion occurs) and on the change in the superfluid velocity (in regions in which vortex motion occurs during the glitch). (3) It leads to relaxation times proportional to the internal temperature of the star, so that timing observations can provide a unique probe of the internal stellar temperature. This internal temperature can then be used to obtain an estimate of the surface temperature.

Thermally activated vortex creep leads to an internal dissipation of the stored rotational energy of the neutron star. The magnitude of the continuous energy dissipation resulting from creep follows directly from the superfluid equations of motion, and is

$$\dot{E}_{\text{diss}} \equiv |\dot{\Omega}_c| \int_p dI_p \omega_{\infty} \equiv |\dot{\Omega}_c| I_p \bar{\omega}_{\text{cr}} \quad (19)$$

where I_p is the moment of inertia of the entire set of pinning layers, $\bar{\omega}_{\text{cr}}$ is an average through the pinning layers and ω_{∞} is the steady-state value of $\Omega - \Omega_c$. This dissipation is the principal source of heat for old pulsars which have radiated away their original heat content.

Vela pulsar glitches

Alpar *et al.*¹⁹ have shown that the observed complex post-glitch behaviour of the Vela pulsar following the four glitches which took place in the decade 1969–79 (ref. 30) finds a straightforward explanation in terms of vortex creep theory. They fit the observational data of Downs with the following expression for $\dot{\Omega}_c(t)$,

$$I_c \dot{\Omega}_c(t) = N_{\text{ext}} - N_1 - N_2 - N_A \quad (20)$$

Here N_1 and N_2 are specified by equations (16) and (17), with relaxation times τ_1 and τ_2 . These torques represent an average response of two distinct pinning layers through which no vortices move. A third region, through which vortices move, is represented to a first approximation by

$$N_A \cong \dot{\Omega}_\infty \frac{I_A t}{t_{OB}} \quad (21)$$

(ref. 19), where I_A is the moment of inertia of the superfluid in pinning layers (where vortices unpin and repin) and t_{OB} is the post-glitch delay time given by equation (18) on taking X_i to be the total number of vortices that unpin. A choice of $\tau_1 = 3$ days and $\tau_2 = 60$ days, for each of the four glitches, gives an excellent fit to the observational data (Fig. 3). Alpar *et al.*¹⁹ conclude that the two distinct pinning layers correspond to weak pinning and superweak pinning respectively, characterized by $\langle \alpha^{SW} \rangle \sim 2 \times 10^{-17}$ days $K^{-1} s^{-2}$ and $\langle \alpha^W \rangle \sim 4 \times 10^{-16}$ days $K^{-1} s^{-2}$ (see equation (15)). Within each pinning region the actual pinning parameters may vary either way by factors of ~ 2 . The size of the inertial moments of these pinning regions ($\sim 10^{-2} I$) is physically reasonable, in terms of current neutron-star models.

From the observed relaxation times, Alpar *et al.*¹⁹ estimate the internal temperature of the Vela pulsar to be $(1.5 \pm 1) \times 10^7$ K. The quoted errors do not refer to a particular calculation, but rather reflect an estimate of the degree of uncertainty introduced by lack of knowledge of various pinning parameters.

Crab pulsar

Alpar, Nandkumar and Pines find, surprisingly, that the post-glitch behaviour of the Crab pulsar following its two large glitches in 1969 and 1975 could be fitted (see Fig. 3b) with the extreme (Occam's razor) hypothesis that both the internal structure and the vortex pinning sites in the Crab pulsar are essentially the same as those hypothesized for the Vela pulsar, so that the only significant physical differences between the two pulsars reside in the spin down rate and internal temperature⁴⁴. In this fashion they deduce an internal temperature for the Crab pulsar of 3.8×10^8 K, with, again, an uncertainty either way of a factor of ~ 2 . The surface temperatures of the Vela and Crab pulsars can be estimated using the scaling laws obtained by Gudmundsson *et al.*⁴⁸. For a $1.4 M_\odot$ neutron star with a reasonably stiff equation of state, one finds the surface temperatures observed at infinity are $T_{\infty, \text{Vela}} \approx 3 \times 10^5$ K and $T_{\infty, \text{Crab}} \approx 1.6 \times 10^6$ K. For both pulsars, these surface temperatures are well below the upper limits obtained by the Einstein observatory^{49,50}. Comparing our estimates with calculations on neutron star cooling and evolution (see ref. 51 for recent review), one finds that an internal temperature ratio of 25 is not consistent with the standard cooling calculations, which, for a given Crab temperature would predict a considerably hotter Vela pulsar. Therefore, these results may provide the first indication that non-standard, more rapid cooling scenarios (perhaps reflecting the presence of a pion condensate, if a pion condensate indeed leads to sufficiently rapid cooling) are in operation in neutron stars.

Old pulsars

Consider now pulsars which are old enough to have radiated away their original heat content so that their dominant source of heat is due to internal dissipation by vortex creep. On equating the rate of dissipation $\dot{E}_{\text{diss}}$ (equation (19)) to the stellar luminosity, one finds a limiting black-body surface temperature of

$$T_s \approx 1.1 \times 10^5 [I_{p,43} \bar{\omega}_{cr} \dot{\Omega}_{-14} / R_6^2]^{1/4} \quad (22)$$

From the upper limit to the surface luminosity of the Vela pulsar in the soft X-ray region⁴⁹ and the Einstein detection of PSR1929+10 (ref. 52), we have estimated $I_{p,43} \bar{\omega}_{cr}$ to be ≤ 1 (ref. 53). For such a pulsar, the thermal flux given by equation (22) is small enough for the internal temperature to be close to the surface temperature. If all old pulsars possess essentially the same pinning layers as the Crab and Vela pulsars, they would have temperatures determined by internal dissipation according

to equation (22) and would possess creep relaxation times that depend only on their spin-down rate, and are

$$\tau_{sw}^{\text{old}} \sim 220 |\dot{\Omega}_{-14}|^{-3/4} \text{ days} \quad (23)$$

$$\tau_w^{\text{old}} \sim 4.4 \times 10^3 |\dot{\Omega}_{-14}|^{-3/4} \text{ days} \quad (24)$$

for the superweak and weak pinning layers, respectively.

Among the old pulsars which have exhibited glitches, post-glitch timing observations are available only for PSR0525+21. Alpar, Nandkumar and Pines⁴⁴ (Fig. 3c) find that vortex creep theory provides a good fit to the observed post-glitch behaviour of this pulsar with a 'fast' relaxation time of ~ 150 days and is consistent with the additional presence of a much slower 3,000-day relaxation process. The vortex creep theory prediction of τ for PSR0525+21 is, from equation (24), some 140 days, in excellent agreement with this fit. Thus vortex creep theory can explain the post-glitch behaviour of both very young and very old pulsars. What is striking is that it does so without involving new parameters for each glitch or pulsar.

Superfluidity in millisecond pulsars

The recently discovered millisecond pulsars^{54,55} have much larger angular velocities, Ω , and much lower inferred dipolar fields, B_s , than 'ordinary' pulsars. Neither of these changes alters the superfluid physics of the neutron star⁵³. The search for observational consequences of superfluidity in millisecond pulsars thus becomes a search for pinned crustal superfluid phenomena: (1) the internal generation of heat by vortex creep; (2) macroglitches, such as those observed in the Vela pulsar, produced by the simultaneous unpinning of a large group of vortex lines; (3) post-glitch timing behaviour associated with the gradual recoupling of the pinned superfluid; (4) timing noise.

The first three phenomena scale with $\dot{\Omega}_c$, whereas the value of $|\dot{\Omega}_c| (= 2.9 \times 10^{-13} \text{ rad s}^{-2})$ for the millisecond pulsar PSR1937+21 resembles that of ordinary pulsars. Our results are not encouraging for would-be observers of glitches in millisecond pulsars; although our estimated glitch intervals ($\sim 10^2$ to $\sim 10^4$ yr) are comparable with those found in the conventional pulsar sample⁵⁵, the present sample of millisecond pulsars is too small to make it likely that a substantial glitch in one of these will be observed in the near future.

The timing from millisecond pulsars is of great importance in view of the potential practical applications of the excellent timing accuracy⁵⁶ of PSR1937+21. To estimate the contribution of the unpinning events to timing residuals, we take the rate of unpinning events to be $R \sim |\dot{\Omega}_c| / \delta\Omega_{cr}$ where $\delta\Omega_{cr}$ is a critical lag between superfluid and crust rotation rates needed to trigger a typical unpinning event, and assume that the unpinning event involves an angular momentum transfer of order $\Delta L = I_c \Delta\Omega_c \equiv I_i \delta\Omega_{cr}$ where I_i is the moment of inertia of the region through which vortices move in a typical unpinning event. The noise strength S then scales with $\dot{\Omega}_c$:

$$S \equiv R \langle \Delta\Omega_c^2 \rangle \sim |\dot{\Omega}_c| \left(\frac{I}{I_i} \right)^2 \delta\Omega_{cr} \quad (25)$$

There is some evidence⁵⁷ for a correlation of S with $|\dot{\Omega}_c|$. On scaling from the observations of white $\dot{\Omega}_c$ noise in the noisiest source, the Crab pulsar, and expressing the result as an r.m.s. timing residual⁵⁸, one has

$$\langle R^2 \rangle_{1937+21, \text{intrinsic}}^{1/2} \sim 4 \times 10^{-2} \mu\text{s} \quad (26)$$

$\sim 5\%$ of the current residual $\approx 0.9 \mu\text{s}$ (ref. 56). Thus, one can be (cautiously) optimistic that the intrinsic contributions to the noise in PSR1937+21 are quite small. Note, however, that if the events are triggered by a critical value $(\delta\Omega/\Omega)_{cr}$ instead of $\delta\Omega_{cr}$, $S \propto \Omega_c \dot{\Omega}_c$ and scaling from the Crab pulsar yields an estimate of intrinsic noise comparable with the observed residuals in PSR1937+21.

Accreting X-ray sources

We have concentrated on effects of superfluidity in radio pulsars, where the external torque can be taken to be constant on time-

scales of the observed variations in timing behaviour. In the new picture of neutron-star superfluids that we have presented, only $\sim 10^{-2}$ of the superfluid interior, the pinned superfluid in the crust, produces observable deviations from rigid-body behaviour, and the resulting Ω_c variations are of the same order. Such variations are clearly unimportant in comparison with the observed variations in the accretion torque in systems like Her X-1 or Vela X-1 (ref. 59 and F. K. Lamb *et al.*, in preparation), so that the superfluid interior is negligible as a source of timing noise in accreting X-ray pulsars.

The dynamics of the pinned superfluid does, however, place severe constraints on precession (wobble) frequencies and damping times, which in effect rule out neutron-star precession as the explanation of various quasi-regular timescales observed in X-ray binaries (for example, the 35-day cycle in Her X-1). First, Shaham⁶⁰ has pointed out that the presence of pinned superfluid leads to wobble frequencies Ω_w given by $\Omega_w/\Omega \sim I_p/I \sim 10^{-2}$, where Ω is the rotation frequency and I the total moment of inertia. Second, precession of the crust will be damped by its coupling to a fluid interior. The damping time is $\tau_d \approx \tau\Omega/\Omega_w$ (ref. 61) where τ is the dynamical coupling time, and is thus of the order of 5 days for a neutron star with a 10-s rotation period.

Concluding remarks

We have described in this article the key role which synoptic observations and detailed analysis of the timing data of individual pulsars has played in bringing about a substantive change in perspective on post-glitch relaxation and the astrophysical evidence for neutron superfluidity. Without observational data of this quality there would have been little incentive to

develop a detailed theory of pinned vortex motion. Future careful observations of Crab, Vela and other pulsars will be of assistance in pointing out directions in which the theory can be improved and refined.

Much work has been done on neutron-star superfluidity which, unfortunately, there has not been space to discuss here. This includes work on the nature of pion condensates in neutron stars^{11,12,62,63} neutron-star cooling⁵¹, the dynamics of moving vortex lines⁶⁴, as well as work in progress (M.A.A., R. Nandkumar and D.P., in preparation) on information concerning vortex creep theory which may be provided by the recent analysis (ref. 65 and P. Boynton and J. Deeter, in preparation) of timing noise observations to 23 pulsars in addition to Crab and Vela.

It could be argued that, although a convincing case has been made for crustal neutron superfluidity, there is little observational evidence for core neutron superfluidity, as the calculated rapid coupling of the core superfluid to the crust rules out post-glitch behaviour as a way to observe core superfluids. Better observations of pulsar surface temperatures and improved theories of neutron-star cooling may offer the best prospects for determining the role played by core superfluids. As we have noted, rapid cooling of pulsars during the epoch from $\sim 10^3$ to 10^4 yr is required by the internal temperatures inferred above for the Crab and Vela pulsars. This may provide a cogent theoretical and observational argument in favour of pion condensation, but more work will need to be done before any definitive conclusions can be reached.

We thank our colleagues in Urbana for many helpful discussions, Radha Nandkumar for a numerical computation and Fred Lamb and Noriaki Shibasaki for critical reading of the manuscript. This work has been supported in part by NSF grant PHY80-25605 and NASA grant NSG-7653.

1. Migdal, A. B. *Nucl. Phys.* **13**, 655-674 (1959).
2. Ginzburg, V. L. & Kirzhnits, D. A. *Sov. Phys. JETP* **20**, 136-1348 (1965).
3. Baade, W. & Zwicky, F. *Phys. Rev.* **45**, 138 (1933).
4. Oppenheimer, J. R. & Volkoff, G. M. *Phys. Rev.* **55**, 374 (1939).
5. Pacini, F. *Nature* **216**, 567-568 (1967).
6. Pines, D. *Proc. 12th int. Conf. Low Temp. Phys.*, 7-17 (Kaigaku, Tokyo, 1971).
7. Pandharipande, V. R., Pines, D. & Smith, R. A. *Astrophys. J.* **208**, 550-566 (1976).
8. Pines, D. *J. Phys.* **41**, C2-111-124 (1980).
9. Ruderman, M. *Nature* **223**, 597-598 (1969).
10. Baym, G. & Pethick, C. J. *A. Rev. Astr. Astrophys.* **17**, 415-443 (1979).
11. Takatsuka, T. in *Recent Progress in Many-Body Theories* (eds Zabolitzky, J. G., de Llano, M., Fortes, M. & Clark, J. W.) 453-461 (Springer, Berlin, 1981).
12. Tamagaki, R. in *Recent Progress in Many-Body Theories* (eds Zabolitzky, J. G., de Llano, M., Fortes, M. & Clark, J. W.) 444-452 (Springer, Berlin, 1981).
13. Hoffberg, M., Glassgold, A. E., Richardson, R. W. & Ruderman, M. *Phys. Rev. Lett.* **24**, 775-777 (1970).
14. Yang, C.-H. & Clark, J. W. *Nucl. Phys.* **A174**, 49-62 (1971).
15. Takatsuka, T. *Prog. theor. Phys.* **71**, 1432-1435 (1984).
16. Takatsuka, T. *Prog. theor. Phys.* **48**, 1517-1533 (1972).
17. Alpar, M. A. *Astrophys. J.* **213**, 527-530 (1977).
18. Anderson, P. W., Alpar, M. A., Pines, D. & Shaham, J. *Phil. Mag.* **A45**, 227-238 (1982).
19. Alpar, M. A., Anderson, P. W., Pines, D. & Shaham, J. *Astrophys. J.* **278**, 791-805 (1984).
20. Negele, J. W. & Vautherin, D. *Nucl. Phys.* **A207**, 298-320 (1973).
21. Chao, N.-C., Clark, J. & Yang, C.-H. *Nucl. Phys.* **A179**, 320-332 (1972).
22. Baym, G., Pethick, C. J. & Pines, D. *Nature* **224**, 673-675 (1969).
23. Easson, I. *Astrophys. J.* **228**, 257-267 (1979).
24. Baym, G., Lamb, D. Q. & Lamb, F. K. *Astrophys. J.* **208**, 829-835 (1976).
25. Baym, G., Pethick, C. J., Pines, D. & Ruderman, M. *Nature* **224**, 872-874 (1969).
26. Sauls, J. A., Stein, D. & Serene, J. W. *Phys. Rev. D* **25**, 967-975 (1982).
27. Alpar, M. A., Langer, S. & Sauls, J. A. *Astrophys. J.* **282**, 533-541 (1984).
28. Lohsen, E. *Astr. Astrophys. Suppl.* **44**, 1-14 (1981).
29. Downs, G. S. *Astrophys. J. Lett.* **257**, L67-70 (1982).
30. Downs, G. S. *Astrophys. J.* **249**, 687-697 (1981).
31. McCulloch, P. M., Hamilton, P. A., Royle, G. W. R. & Manchester, R. N. *Nature* **302**, 319-321 (1983).
32. Hamilton, P. A., McCulloch, P. M. & Royle, G. W. R. *IAU Circ. No.* 3729 (1982).
33. Manchester, R. N., Newton, L. M., Hamilton, P. A. & Goss, W. M. *Mon. Not. R. astr. Soc.* **202**, 269-285 (1983).
34. Newton, L. M., Manchester, R. N. & Cooke, D. J. *Mon. Not. R. astr. Soc.* **194**, 841-850 (1981).
35. Backus, P. R., Taylor, J. H. & Damashek, M. *Astrophys. J.* **255**, L63-67 (1982).
36. Alpar, M. A. & Ho, C. *Mon. Not. R. astr. Soc.* **204**, 655-667 (1983).
37. Boynton, P. E. *Proc. IAU Symp.* No. 95, 279-290 (1981).
38. Demianski, M. & Proszynski, M. *Mon. Not. R. astr. Soc.* **202**, 437-451 (1983).
39. Baym, G. & Pines, D. *Ann. Phys.* **66**, 816-835 (1971).
40. Anderson, P. W. & Itoh, N. *Nature* **256**, 25-27 (1975).
41. Pines, D., Shaham, J., Alpar, M. A. & Anderson, P. W. *Prog. theor. Phys. Suppl.* **69**, 376-396 (1980).
42. Shaham, J. & Pines, D. *Proc. 1981 Annual Meeting, European Physical Society* (1982).
43. Alpar, M. A., Anderson, P. W., Pines, D. & Shaham, J. *Astrophys. J.* **276**, 325-334 (1984).
44. Alpar, M. A., Nandkumar, R. & Pines, D. *Astrophys. J.* **288**, 191-195 (1985).
45. Feibelman, P. J. *Phys. Rev. D* **4**, 1589-1597 (1971).
46. Anderson, P. W. *Phys. Rev. Lett.* **9**, 309-311 (1963).
47. Anderson, P. W. & Kim, Y. B. *Rev. mod. Phys.* **36**, 39-43 (1964).
48. Gudmundsson, E. H., Pethick, C. J. & Epstein, R. I. *Astrophys. J. Lett.* **259**, L19-23 (1982).
49. Harnden, F. R. Jr, Grant, P. D., Seward, F. D. & Kahn, S. M. preprint (1985).
50. Harnden, F. R. Jr & Seward, F. D. *Astrophys. J.* **283**, 279-285 (1984).
51. Tsuruta, S. *Physics Reports* (in the press).
52. Helfand, D. *Proc. IAU Symp.* No. 101, 471-486 (1982).
53. Pines, D. & Alpar, M. A. in *Proc. NRAO Workshop No. 8*, (eds Reynolds, S. P. & Stinebring, D. R.) 161-172 (NRAO, 1984).
54. Backer, D. C., Kulkarni, S. R., Heiles, C., Davis, M. M. & Goss, W. M. *Nature* **300**, 615-618 (1982).
55. Boriakoff, V., Buccheri, R. & Fauci, F. *Nature* **304**, 417-418 (1983).
56. Davis, M. M., Taylor, J. H., Weisberg, J. M. & Backer, D. C. *Astrophys. J. Lett.* (in the press).
57. Cordes, J. M. & Helfand, D. J. *Astrophys. J.* **239**, 640-650 (1980).
58. Blandford, R., Narayan, R. & Romani, R. W. preprint (1984).
59. Boynton, P. E. *et al. Astrophys. J. Lett.* **283**, L53-56 (1984).
60. Shaham, J. *Astrophys. J.* **214**, 251-260 (1977).
61. Bondi, H. & Gold, T. *Mon. Not. R. astr. Soc.* **115**, 41-46 (1955).
62. Tatsumi, T. *Prog. theor. Phys.* **69**, 1137-1145 (1983).
63. Takatsuka, T. *Prog. theor. Phys.* **72**, 252-265 (1984).
64. Shaham, J. *J. Phys.* **41**, C2-9-23 (1980).
65. Cordes, J. & Downs, G. S. *NRAO preprint* (1984).

Punctiform initiation of seafloor spreading in the Red Sea during transition from a continental to an oceanic rift

Enrico Bonatti

Lamont-Doherty Geological Observatory, Columbia University, Palisades, New York 10964, USA

Transition from a continental to an oceanic rift occurs in the Red Sea by initial emplacement of oceanic crust in regularly spaced 'hot points', which serve as nuclei for axial propagation into segments of oceanic crust accretion and for initiation of seafloor spreading. Hot points develop above upwelling mantle diapirs caused by density/viscosity inversion beneath continental and oceanic rifts.

AN important goal in earth science is to understand how the transition from a continental to an oceanic rift occurs in time and space, which will further our knowledge of oceanic Atlantic-type passive margins. This transition, which occurred along the proto-Atlantic rift 160–90 Myr ago, can be observed today in the northern Red Sea. Based on morphotectonic, magnetometric, seismic-reflection and petrological data from this region, I will attempt to demonstrate that: (1) injection of oceanic crust takes place initially in punctiform hot areas axially spaced ~50 km apart; (2) activation of hot points shows a northward time progression; and (3) each hot point serves as a nucleus for axial propagation of the zone of oceanic crust accretion into linear segments of spreading. I will also explain the punctiform initiation of seafloor spreading in terms of inferred patterns of upper mantle upwelling beneath the Red Sea, and suggest that the same model can be applied to the initial opening of the Atlantic rift.

Plate-tectonic reconstructions of the Red Sea/Gulf of Aden region^{1–5} suggest that, although emplacement of oceanic crust between Arabia and Africa initiated in the Gulf of Aden as early as ~10 Myr ago⁶, it started more recently in the Red Sea⁷. An axial rift valley with strong magnetic signature is present almost continuously in the southern Red Sea between 16° and 19° N. Vine-Matthews-type axial magnetic anomalies indicate initiation of seafloor spreading by injection of basalt crust in a narrow axial zone ~5 Myr BP in a region around 17° N, but more recently to the north and south of this region⁸, in a pattern which suggests axial rift propagation⁹. Moving northward, the axial rift valley and associated magnetic signature become discontinuous and then disappear completely in the northern Red Sea (Figs 1, 2).

This description assumes that, outside the axial zone with clear-cut high-amplitude Vine-Matthews magnetic anomalies, the Red Sea is floored by thinned, stretched continental crust criss-crossed by diffuse basaltic dyke injections, a view supported by recent geophysical data collected at sea^{7,10–12} and by observations at Zabargad (St John) Island, an uplifted block of Red Sea lithosphere¹³. However, the concept that the whole width of the Red Sea formed by two stages of seafloor spreading, the first of which started as early as ~40 Myr BP, has also been proposed^{14,15}.

The axial zone of the Red Sea can be divided from south to north into the following regions (Fig. 1), based on their different morphotectonic and magnetometric characteristics^{7,8,12,16–18}: (1) rift valley region (17° N–19°30' N), with a relatively straight and continuous axial rift valley and linear magnetic anomalies; (2) multi-deeps region (19°30' N–22° N), with a complex pattern of axial deeps and troughs; (3) transitional region (22° N–24° N), with a few regularly spaced axial trough segments; (4) northern region (north of 24° N), with no axial rift valleys but a few

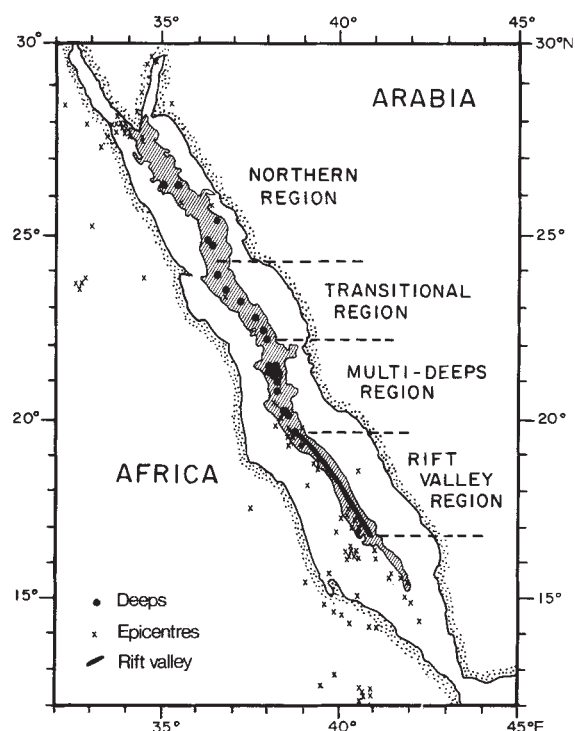


Fig. 1 Outline of the Red Sea, with the axial zone enclosed by the 500-fathoms isobath. The approximate location of the axial rift valley and of isolated deeps and troughs is indicated. Also shown are locations of earthquake epicentres of magnitude >4 recorded from 1953 to 1983. The axial zone has been subdivided into geotectonic regions (see text).

isolated deeps. The model proposed here is based primarily on results obtained from the transitional region.

Red Sea transitional region

Field work in this region (Fig. 2) has shown the existence of segments of axial troughs and deeps with high-amplitude magnetic anomalies, separated by intertrough zones which lack an axial rift valley and high-amplitude magnetic anomalies, and have continuous, thick sediment cover across the axis^{11,16–18}.

The transitional region axial trough segments, named (from south to north) Thetis, Nereus, Bannock and Vema^{11,16}, become shorter, narrower and shallower moving from south to north (Fig. 2a, b). The Thetis and Nereus segments display morphotectonic features typical of rift valleys of oceanic slow-spreading axial zones such as the Mid-Atlantic Ridge (MAR). They have

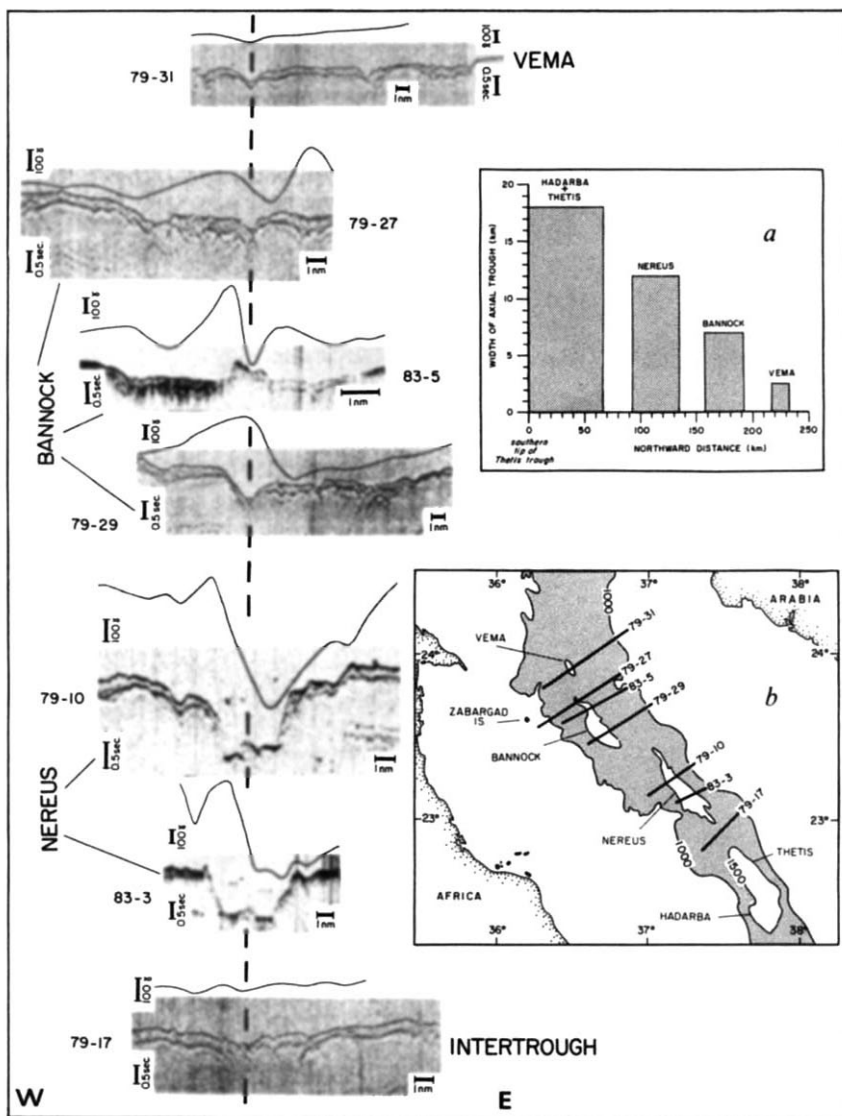


Fig. 2 Outline of the Red Sea transitional region where the axial troughs and deeps (Thetis-Hadarba, Nereus, Bannock and Vema) are shown enclosed by the 1,500-m isobath (insert *a*). A series of seismic reflection profiles taken across the Red Sea axis (locations indicated in insert *b*) is shown on the left of the figure, together with magnetic anomalies along the same profiles. The vertical broken line indicates the approximate axis of the Red Sea. Insert *b* shows the width and length of the Thetis-Hadarba, Nereus, Bannock and Vema axial trough and deeps, defined by the 1,500-m isobath. Length is estimated from the distance along the Red Sea axis taken from the southern end of the Thetis-Hadarba trough.

steep walls formed by inward-facing normal faults¹⁸ and a neovolcanic zone floored by young basalts with mid-ocean ridge basalt (MORB) affinity¹¹. A trough segment to the south of Thetis, termed Hadarba Deep¹⁶, merges with Thetis although the axes of the two segments are offset laterally by ~5 km (Fig. 2b). High-amplitude magnetic anomalies are present at Thetis¹⁵, although its scant coverage does not allow an accurate estimate of the time of initiation of seafloor spreading. Its wider, longer and deeper trough suggests that Thetis is older and more evolved than Nereus. Magnetic anomalies over the Nereus trough suggest the presence of oceanic crust and initiation of seafloor spreading ~1 Myr ago in sections across the northern and southern parts of the trough, but >2 Myr ago in sections from the central part of the trough¹¹. Thus, the Nereus axial trough could be considered a mini-propagating oceanic rift segment where oceanic crust was emplaced initially in a central spot which then served as a nucleus for axial propagation.

The next axial deep segment to the north of Nereus, the Bannock Deep, lacks a steep axial rift valley but shows a gentle graben-like morphology with continuous sediment cover except along one section where a 500-m-high basement protrudes from the sediment (Fig. 2, profile 83-5). This feature is probably a volcanic seamount because fragments of dolerite were recovered from it¹¹. Profiles taken across the axis a few kilometres north and south of the seamount indicate continuous sediment cover and a prominent fault aligned axially with the volcanic seamount (Fig. 2). Narrow high-amplitude magnetic anomalies are associated with the seamount and the axial fault.

The Vema axial deep consists of a relatively gentle depression with continuous sediment cover and no high-amplitude magnetic anomalies, but there is relatively high heat flow¹¹. A consistent pattern of axial high-amplitude magnetic anomalies is absent in the Red Sea north of the Vema Deep, although deeps and anomalies have been found at several isolated sites^{16,18,19}.

Punctiform initiation of seafloor spreading

The observations from the Red Sea transitional region lead to the following interpretations, summarized in part in Fig. 3:

(1) The initial breakthrough of oceanic crust and the beginning of organized seafloor spreading occurs in essentially punctiform hot areas located in axially oriented segments of tensional fissures. The initial breakthrough of oceanic crust would occur through continental crust which has been thinned, stretched and diffusely injected by basaltic dykes^{7,10,11}, although it could occur in ancient, cold oceanic crust if the two-stage opening model^{14,15} is accepted. The igneous body cropping out from the axis of the Bannock graben and the axial fault detected just south and north of this igneous body are examples of recent hot point formation.

(2) Axial propagation from each hot point gives rise to linear spreading segments⁹. This hypothesis is supported by the detection of magnetic anomalies, as old as 2–3 Myr in sections from the centre of the Nereus trough segment, but younger close to the tips of the segment.

(3) A northward progression can be detected in the time of initial activation of the punctiform areas of ocean crust emplace-

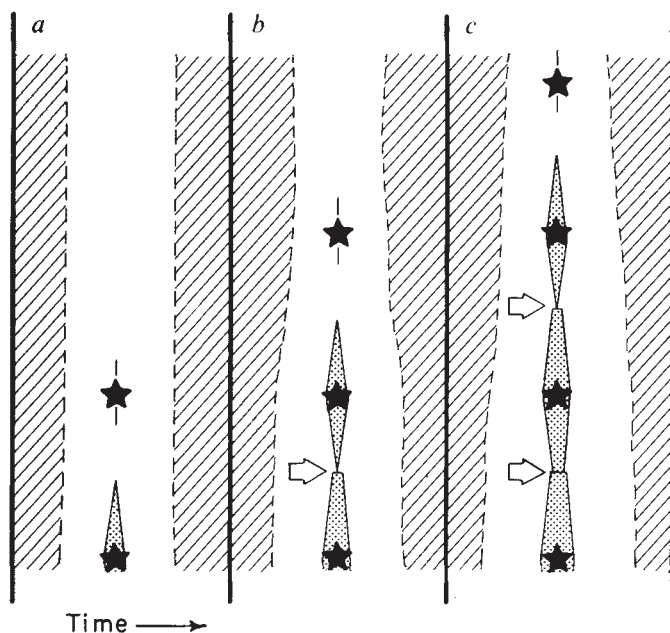


Fig. 3 Schematic model of the initiation of seafloor spreading in the Red Sea transitional region. *a-c*, Different stages in time. Stars indicate sites of hot points where the initial emplacement of oceanic crust occurs. Striped areas, continental; stippled areas, oceanic; open areas, intermediate continental crust stretched, thinned and injected by basaltic dykes. Arrows, discontinuities in the morphotectonic and magnetic axial pattern, predicted by the model.

ments. Although in the southern Red Sea (between 17° N and 18° N), seafloor spreading initiated ~ 5 Myr BP, emplacement of oceanic crust probably started between 2 and 3 Myr BP at Nereus; it started <1 Myr BP at Bannock and has not yet started at Vema. The observation that the Thetis trough is deeper and wider than Nereus suggests that it is more evolved, and is therefore older than Nereus. Relatively high heat flow, as detected within the Vema Deep¹¹, may be a precursor to oceanic crust emplacement and may therefore mark future hot points in the northern Red Sea.

(4) The distance between the centre points of Thetis, Nereus, Bannock and Vema is ~ 50 km (Fig. 2b), suggesting a regular spacing of the areas of upper mantle upwelling and oceanic crust initial emplacement.

(5) Axial propagation from each hot point will eventually lead to joining of the different rift segments. Discontinuities are likely to develop in the axial morphotectonic and magnetic lineations where the tip of rift segments derived from different hot points join (Fig. 3). Moreover, the axial troughs and deeps are not all aligned; some show lateral offsets (Fig. 2). These initial offsets in the segments of oceanic crust emplacement may or may not lead to the formation of initial transform faults.

(6) The regular segmentation of the axis of spreading (Fig. 3) is interrupted where the Red Sea lithosphere is intersected by pre-existing shear zones²⁰, which would prevent normal axial rift propagation and act as 'locked zones' in the sense described by Courtillot⁹.

I now consider how the proposed model, derived from observations in the transitional region, fits available data from other regions of the Red Sea.

Rift Valley region. The evolution of seafloor spreading from a punctiform initiation can be inferred from observations obtained in the southern Red Sea between 17° N and $19^{\circ}30'$ N (Fig. 1). The relatively continuous axial trough and the associated high-amplitude magnetic anomalies imply a steady though asymmetrical seafloor spreading regime for the past few million years, with spreading being about two-thirds faster to the west than to the east^{8,16}.

A detailed investigation of the axial zone at 18° N, including direct submersible observations^{21,22}, has revealed a morpho-

tectonic structure of the axial trough similar to that of the MAR in the FAMOUS area, a neovolcanic zone with basalts similar to MORB, and a spreading rate averaging 1.6 cm yr^{-1} for the past 3–5 Myr. A deep-tow profile (Fig. 4) of the axial zone at $17^{\circ}30'$ N (ref. 23) confirmed a morphotectonic structure similar to that found in the FAMOUS area, and demonstrated asymmetrical spreading ($\approx 5 \text{ mm yr}^{-1}$ to the east, 10 mm yr^{-1} to the west) starting at 4.8 Myr BP. In the context of the proposed model it is of interest that the Rift Valley morphotectonic structures and high-amplitude magnetic anomalies run along the axis in this region for straight segments 30–50-km long, which are stepped laterally by small, 3–10-km transform or non-transform relay zones (Fig. 4) where the magnetic anomalies are weak and irregular¹². One such discontinuity was surveyed in detail at 18° N (ref. 21). The proposed model explains these discontinuities as marking sites where different rift segments meet, each created by longitudinal propagation from an initial hot point (Fig. 3).

Multi-deeps region. This region, located between the more evolved southern Red Sea and the transitional region (Fig. 1), displays a complex geotectonic situation: the straight axial rift valley is replaced by a network of deeps distributed partly in an 'echelon' pattern^{16–18,24}. Some of these deeps are thermally active at present and contain hot brines (for example, Atlantis II, Discovery); some are instead inactive²⁴. This complex geotectonic structure may be determined in part by the presence of a major transverse fracture zone intersecting the Red Sea axis^{17,20,25}, an interpretation supported by the clustering of recent earthquake epicentres in this area (Fig. 1), and by the orientation of magnetic anomalies transversal to the Red Sea axis¹⁷. This is one of three major fracture zones intersecting the Red Sea, the other two being the Zabargad and the Brothers fracture zones, all striking parallel to the Dead Sea–Aqaba fault²⁰.

Northern region. The Red Sea north of the Vema Deep lacks a regular pattern of axial rift valley segments and associated high-amplitude magnetic anomalies^{7,11}, though a few isolated deeps are present, some located off the axis (Fig. 1). Some are associated with magnetic anomalies; that is, Charcot Deep at 25° N, where non-MORB basalts intrude the sediment¹⁹, and Al Wajab Deep at $25^{\circ}20'$ N, where the anomalies are lineated partly in a direction transversal to the Red Sea axis^{15,19}. Some of these deeps may represent small pull-apart basins caused by tectonics related to the fracture zone, such as Al Wajab, at the northeastern end of the Zabargad Fracture Zone, where the Red Sea axial direction shifts by $\approx 40^{\circ}$ (Fig. 1).

Following Cochran⁷, I interpret the northern Red Sea region as floored by thinned and stretched continental crust injected by diffuse basaltic intrusions. Probable examples of such intrusions are basaltic/gabbroic rocks from the Brothers Islets²⁶ and the Charcot Deep¹⁹, showing transitional or alkaline affinity, in contrast to the MORB-like characteristics of basalts from the Nereus Trough¹¹ and from other axial sites further south in the Red Sea²⁷. Thus, punctiform axial injection of oceanic crust and organized seafloor spreading have probably not yet started in the northern region of the Red Sea.

Mantle upwelling beneath the Red Sea

The punctiform model of initiation of seafloor spreading in the Red Sea implies that beneath each initial hot point there is a zone of preferential upwelling of upper mantle material. It is generally agreed that a broad thermal anomaly affects the upper mantle in major rift zones of the Earth. One of the tenets of plate-tectonic theory is that the asthenosphere/lithosphere boundary rises close to the surface along the axis of oceanic rifts²⁸. A rise of the upper-mantle isotherms and of the asthenosphere/lithosphere boundary has been documented for major continental rifts, such as the East African²⁹, Rio Grande^{30,31}, Rhinegraben³² and Baikal^{33,34} rifts. A rise of upper-mantle isotherms and a thinning and stretching of the lithosphere, probably resulting in part from convective heat transfer to its base³⁵, obtains also beneath the Red Sea rift, a view consistent

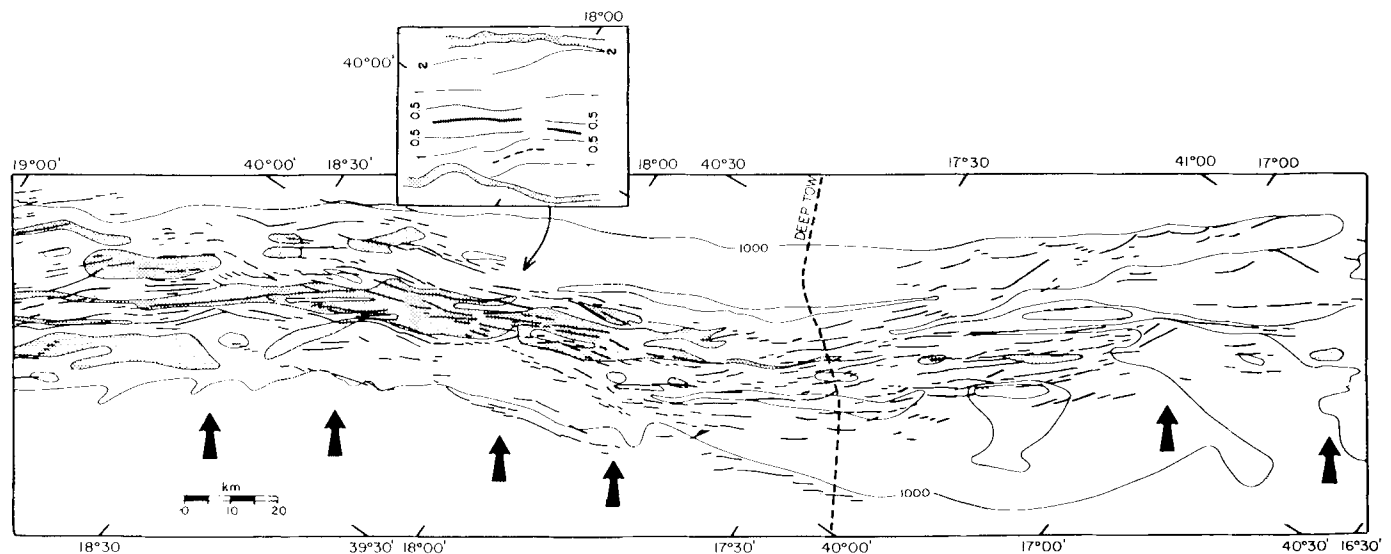


Fig. 4 A portion of the axial zone of the rift valley region of the Red Sea, with simplified bathymetry, from Backer *et al.*¹⁶, and structural lineations obtained by a GLORIA sidescan survey, from Garfunkel *et al.*¹². Arrows indicate discontinuities in the axial morphotectonic lineations observed by Garfunkel *et al.*¹². Insert shows details of one such discontinuity at 18°, simplified from Zonenshajn *et al.*²¹. Isochrons (in Myr), based on magnetic anomalies, are indicated. Thick full line, present axis; dashed thick line, fossil axial segment; shadowed area, outer slopes of rift valley.

with a broad band of high heat flow³⁶ and positive gravity anomaly³⁷ extending across the entire width of the Red Sea.

Continuous upwelling and decompression of upper-mantle material below rift zones leads to partial melting. A low-density/viscosity layer containing a continuously replenished small melt fraction would result, sandwiched between material of higher density and viscosity. The low density of this layer is enhanced by the probability that the upwelling mantle is injected by metasomatic, H₂O- and CO₂-rich fluids, resulting in H₂O-containing phases, such as amphiboles. This mantle metasomatism is important below continental rifts³⁸.

The presence beneath the Red Sea of an upper mantle zone containing a melt fraction and affected by metasomatism is suggested by studies of a mantle-derived spinel lherzolite body uplifted on Zabargad Island (Fig. 2b), probably during the early stages of Red Sea rifting^{13,39}. The lherzolite body contains zones of plagioclase peridotite, probably representing small fractions of melt trapped by lherzolites in the upper mantle, as well as amphibole peridotites produced by reactions of metasomatic fluids with the lherzolite⁴⁰.

The density/viscosity inversion in the upper mantle can be treated as a Raleigh-Taylor-type instability, an approach which has been used to model the distribution of volcanic activity along island arcs⁴¹ and oceanic spreading zones^{42,43}. Theory and experiments^{41,44-47} indicate that the interface between a high density/viscosity upper layer and a layer below with lower density/viscosity develops a sinusoidal instability that may evolve into a pattern of λ -spaced injections of the lower density fluid into the denser layer above. The density contrast affects the velocity of upwelling of the diapirs according to Stoke's law. Their spacing, λ , is related to the ratio of the thicknesses of the upper and lower layer, h_1/h_2 , and the ratios of their viscosities, μ_1/μ_2 (refs 45-47).

In the case of the Red Sea rift, the low-density upper mantle

zone can be assumed to be elongated in shape and of limited width. Provided $h_1 > h_2$, the spacing of the diapirs surging from the lower layer is approximated by^{41,45,46}:

$$\lambda = \frac{2\pi h_2}{2.15} (\mu_1/\mu_2)^{1/3}$$

Assuming μ_1 to be of the order of 10^{20} P^{48,49}, μ_2 can be estimated as being lower by 2-4 orders of magnitude, depending on the degree of partial melting chosen for h_2 and on the temperature and pressure contrast of the two layers. The thickness of h_2 depends on the upper mantle thermal regime of the particular rift being considered. Assuming h_2 as the layer containing melt within the 5-10% range, its thickness beneath the Red Sea axis should be a few kilometres, as estimated from data from Iceland⁵⁰ and mid-ocean ridges^{43,48}. If $\mu_1/\mu_2 = 10^3$ and $h_2 = 2$ km are chosen, a spacing slightly over 50 km is obtained for the upwelling diapirs.

Upwelling of the diapirs would result in an increased degree of melting, with the melt tending to segregate from its parent, collect towards the top of the diapirs and hence be injected into the crust. The attenuated thickness of the Red Sea crust would permit the 'asthenosphere geoid'^{51,52} and hence the melt to reach the sea floor. This mantle-driven pattern leading to the punctiform geometry of initial breakthrough of oceanic crust (Fig. 5), however, is interrupted where the Red Sea lithosphere is intersected and disrupted by pre-existing fracture zones, as in the multi-deeps region and in the Zabargad Fracture Zone.

The northward time progression in the activation of the hot points and in the evolution of seafloor spreading in the Red Sea suggests that the asthenospheric diapirs are superimposed on a broader zone of shallow hot upper mantle which may be moving as a wavefront towards the northern Red Sea (Fig. 5).

Axial propagation from each initial hot point could take place by mechanisms similar to the asthenospheric longitudinal flow

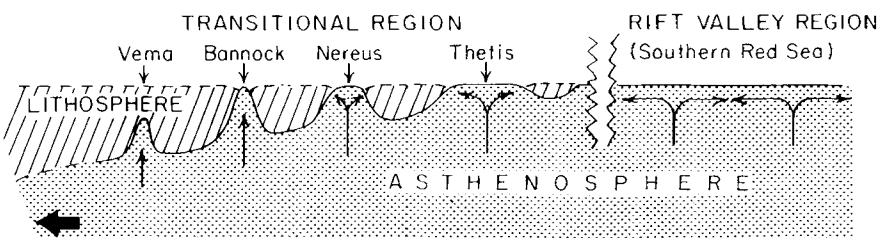


Fig. 5 Schematic and hypothetical cross-section along the axis of the Red Sea, showing regularly spaced asthenospheric diapirs in the transitional region, possibly caused by Raleigh-Taylor instability in the upper mantle, superimposed on a broad wavefront of shallow hot mantle moving northward beneath the Red Sea.

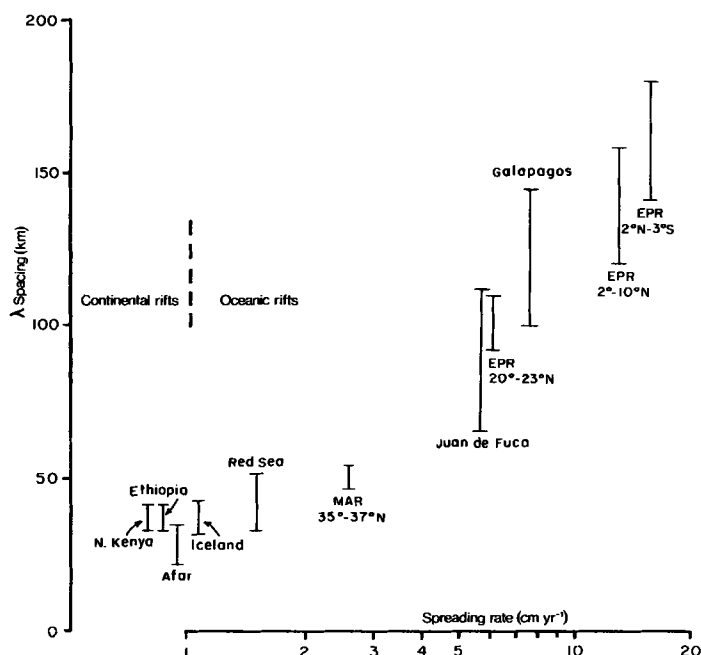


Fig. 6 Spacing of axial volcanism in continental, proto-oceanic and oceanic rifts, versus spreading rate for the oceanic rifts. Data for the East Pacific Rise (EPR), MAR, Galapagos rift and Juan De Fuca Ridge are from Crane⁴³. Data for the Red Sea are from the present paper; for Iceland they include Quaternary and Recent volcanism of the East Volcanic Zone, from Sigurdsson and Sparks⁵⁶; for North Kenya, Ethiopian and Afar rifts they include Quaternary and Recent volcanism, from Mohr and Wood⁵⁷.

described by Courtillot⁹ and Vogt⁵³, and would result in axial spreading segments similar to those 30–50-km long observed in the rift valley region of the southern Red Sea. Interestingly, Mesozoic (155–108 Myr BP) magnetic anomalies from the western North Atlantic reveal that close to the time of opening, the early Atlantic accretionary axis was divided in segments

each ~50 km long with discontinuities of the magnetic anomalies, not necessarily associated with transform offsets, between each segment⁵⁴. Therefore, it seems that the punctiform model of initiation of seafloor spreading proposed for the Red Sea can also be applied to the early Atlantic rift.

Evolution from continental to oceanic rift

Structural studies of peridotites from ophiolite complexes suggest vertical flow of upper-mantle material beneath oceanic ridges, with upwelling zones spaced 50–100 km apart⁵⁵. Crane⁴³ noted that the spacing between areas of upwelling and igneous activity in oceanic zones of crustal accretion is related to spreading rate, that is, ~150 km along the fast East Pacific Rise, ~100 km along the intermediate Galapagos rift and Juan de Fuca Ridge, and ~50 km along the slow MAR. A regular spacing of ~40 km has been noted for active volcanoes in the eastern volcanic zone of Iceland, part of the MAR⁵⁶. The concept of regularly spaced volcanism can be extended to continental rifts: a spacing of ~40 km has been reported for Pleistocene and active volcanoes of the northern Kenya and Ethiopian rifts, and of 30–10 km for those of the Afar rift⁵⁷. The increase in spacing of hot points from continental to proto-oceanic, to slow, intermediate and fast oceanic rifts (Fig. 6) is paralleled by an inferred uplift of sub-rift isotherms and asthenosphere/lithosphere boundary, an increase in the degree of upper-mantle melting and a change in melt chemistry from alkaline to transitional to tholeiitic and to MORB. The first factor above, causing an increase in the thickness of the zone of partial melting from continental to slow to fast-spreading rifts^{43,48}, would determine a corresponding increase in the spacing of asthenospheric diapirs and volcanism. Thus, the transition from continental to oceanic rift as seen in the central Red Sea could be considered an important step in a continuum of rift processes, related ultimately to the evolution of mantle thermal waves.

The field work which led to this paper was sponsored by the Italian Research Council (CNR). I thank the officers and crew of RV *Bannock*; D. Breger, P. Colantoni, K. Crane and M. Taviani for collaboration; and J. Cochran for help in modelling the magnetic anomalies. The research was supported by NSF grant O.C.E. 83-15933 and by ONR contract N00014-8-C-0098. Contribution no. 3833 of the Lamont-Doherty Geological Observatory.

Received 5 December 1984; accepted 5 May 1985.

- McKenzie, D. P. *Earth planet. Sci. Lett.* **40**, 25–32 (1978).
- Freund, R. *Nature* **228**, 453–455 (1970).
- Girdler, R. W. & Darracot, B. W. *Geophysics* **2**, 131–138 (1972).
- Richardson, E. S. & Harrison, C. G. A. *Earth planet. Sci. Lett.* **30**, 135–142 (1976).
- LePichon, X. & Francheteau, J. *Tectonophysics* **46**, 369–406 (1978).
- Cochran, J. R. *Geophys. J. R. astr. Soc.* **68**, 171 (1982).
- Cochran, J. R. *Bull. Am. Ass. Petrol. Geol.* **67**, 41–69 (1983).
- Roeser, H. A. *Geol. Jb.* **13**, 131–153 (1975).
- Courtillot, V. E. *Tectonics* **1**, 239–250 (1982).
- Coleman, R. G. *Proc. 27th Int. geol. Congr., Moscow S-06-2-1* **23**, 93–121 (1984).
- Bonatti, E., Colantoni, P., Della Vedova, B. & Taviani, M. *Oceanologica acta* **7**, 385–398 (1984).
- Garfunkel, Z., Ginzburg, A. & Searle, R. C. *Ann. Geophysica* (in the press).
- Bonatti, E. *et al. J. geol. Soc.* **140**, 677–690 (1983).
- Girdler, R. W. & Styles, P. *Nature* **247**, 7–11 (1974).
- Hall, S. A. *Saudi Arabian Proj. Rep.* **275** (US Geol. Surv. 1979).
- Backer, H., Lange, K. & Richter, H. *Geol. Jb.* **D13**, 79–123 (1975).
- Searle, R. C. & Ross, D. A. *Geophys. J. R. astr. Soc.* **43**, 555–572 (1975).
- Pautot, G. *Oceanologica Acta* **6**, 235–244 (1983).
- Pautot, G., Guennoc, P., Coutelle, A. & Lyberis, N. *Nature* **310**, 133–136 (1984).
- Crane, K. & Bonatti, E. *Tectonophysics* (submitted).
- Zonenshajn, L. P., Monin, A. S. & Sorokhtin, G. *Geotektonika* **2**, 3–22 (1981).
- Juteau, T. *et al. Bull. Centres. Rech. Explor. Prod. Elf-Aquitaine* **7**, 217–231 (1983).
- Miller, S. P., MacDonald, K. C. & Lonsdale, P. F. *J. geophys. Res.* (in the press).
- Degens, E. T. & Ross, D. A. (eds) *Hot Brines and Recent Heavy Metal Deposits in the Red Sea* (Springer, Berlin, 1969).
- Garson, M. S. & Krs, M. *Bull. geol. Soc. Am.* **87**, 169–181 (1976).
- Taviani, M., Bonatti, E., Colantoni, P. & Rossi, P. L. *Boll. geol. Soc. Ital.* (in the press).
- Schilling, J. G. *Science* **165**, 1357–1360 (1969).
- Slater, J. G. & Francheteau, J. *Geophys. J. R. astr. Soc.* **20**, 509–542 (1970).
- Baker, B. H. & Wohlenberg, J. *Nature* **229**, 538–542 (1971).

- Bridwell, R. J. in *Tectonics and Geophysics of Continents. Rifts* (eds Ramberg, H. & Neumann, E. R.) **73–80** (Reidel, Boston, 1978).
- Parker, E. C., Davis, P. M., Evans, J. R., Iyer, H. M. & Olsen, K. H. *Nature* **312**, 354–356 (1984).
- Edel, J. B., Fuchs, K., Gelbke, C. & Prodehl, C. *J. geophys. Res.* **41**, 333–345 (1975).
- Logatchev, N. A., Rogozhina, V. A., Solonenko, V. P. & Zorin, Y. A. in *Tectonics and Geophysics of Continental Rifts* (eds Ramberg, H. & Neumann, E. R.) **49–62** (Reidel, Boston, 1978).
- Zorin, Y. A. *Tectonophysics* **73**, 91–104 (1981).
- Spohn, T. & Schubert, G. *J. geophys. Res.* **87**, 4669–4681 (1982).
- Girdler, R. W. & Evans, T. R. *Geophys. J. R. astr. Soc.* **51**, 245–251 (1977).
- Allan, T. D. *Phil. Trans. R. Soc. A267*, 153–180 (1970).
- Bailey, D. K. *Nature* **296**, 525–530 (1982).
- Bonatti, E., Ottonello, G. & Hamlyn, P. R. *Geology* **9**, 474–479 (1981).
- Bonatti, E., Ottonello, G. & Hamlyn, P. R. *J. geophys. Res.* (in the press).
- Marsh, B. D. *J. Geol.* **87**, 687–713 (1979).
- Whitehead, J. A., Dick, H. J. B. & Shouten, H. *Nature* **312**, 146–148 (1984).
- Crane, K. *Earth planet. Sci. Lett.* **72**, 405–414 (1985).
- Dances, Z. F. *Geophysics* **29**, 414–424 (1964).
- Biot, M. A. & Ode, H. *Geophysics* **30**, 213–227 (1965).
- Seelig, F. *Geophysics* **30**, 633–634 (1965).
- Ramberg, H. *Phys. Earth planet. Inter.* **5**, 45–60 (1972).
- Lewis, B. T. R. *J. geophys. Res.* **86**, 4868–4880 (1981).
- Rabinowicz, M., Nicolas, A. & Vigneresse, J. L. *Earth planet. Sci. Lett.* **67**, 97–108 (1984).
- Thayer, R. E., Bjornsson, A., Alvarez, L. & Hermance, J. F. *Geophys. J. R. astr. Soc.* **65**, 423–442 (1981).
- Turcotte, D. L., Haxby, W. F. & Ockendon, J. R. in *Island Arcs, Deep Sea Trenches and Back-Arc Basins* (eds Talwani, M. & Pitman, W. C.) **63–69** (Am. Geophys. Union, 1977).
- LePichon, X. & Sibuet, J. C. *J. geophys. Res.* **86**, 3708–3720 (1981).
- Vogt, P. R. *Earth planet. Sci. Lett.* **29**, 309–325 (1976).
- Shouten, H. & Klitgord, K. D. *Earth planet. Sci. Lett.* **59**, 255–266 (1982).
- Nicolas, A. & Violette, J. F. *Tectonophysics* **81**, 319–339 (1982).
- Sigurdsson, H. & Sparks, S. R. *Nature* **274**, 126–130 (1978).
- Mohr, P. A. & Wood, C. A. *Earth planet. Sci. Lett.* **33**, 126–144 (1976).

Correcting an immune-response deficiency by creating E_α gene transgenic mice

Marianne Le Meur, Pierre Gerlinger, Christophe Benoist & Diane Mathis

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg Cedex, France

We have introduced a cloned E_α^k gene into embryos of mice incapable of expressing their endogenous E_α genes. The transcription of E_α^k was accurate and tissue-specific in the resulting transgenic lines and, in macrophages, E_α^k transcription could be induced by γ -interferon. The injected gene conferred on the genetically deficient host strain the ability to mount an immune response to a synthetic polymer.

THE immune response to a foreign antigen involves the highly regulated expression of several specific gene sets. Mounting a successful antibody response to antigen challenge depends on mobilizing a repertoire of lymphoid cells: macrophages and dendritic cells that process and present antigen, B cells that effect the actual production of immunoglobulins and T cells that modulate the B-cell response.

A group of loci important in controlling the murine immune response have been mapped to the major histocompatibility complex (MHC) on chromosome 17 (see refs 1–5 for reviews). The 'immune-response' genes encode the Ia antigens, glycoproteins that occur at the cell surface as heterodimeric complexes, the A complex (A_α and A_β) and the E complex (E_α and E_β). Ia molecules are limited in distribution primarily to B cells and antigen-presenting cells (APC) including macrophages; certain immunological 'activators' such as γ -interferon (IFN- γ) and B-cell growth factor increase their expression. Immune-response genes are characterized by an extensive polymorphism, such that any two mouse strains have Ia chains that differ by 1–10% of their amino acids. Certain strains lack the E complex entirely.

The polymorphism of the Ia antigens seems to be the key to their function in the immune response. T cells can only 'see' foreign antigen in the context of an Ia complex and respond only when an APC 'presents' antigen in conjunction with an Ia molecule of the appropriate haplotype. Some antigens are presented in association with the E complex, others with the A; hence particular antigens are said to be E- or A-restricted. T cells activated in response to antigenic stimulation are essential for the B-cell response; this T-B interaction is also subject to 'restriction' by Ia molecules of the appropriate haplotype. The net result of joint antigen-Ia recognition by T cells is that when challenged with a given foreign antigen, antibody is produced depending on the particular Ia alleles expressed.

Recently, the genes for all four Ia chains have been cloned and multiple alleles of each sequenced². Experiments are being performed to localize the regions of each chain involved in restriction, and to define the mechanisms involved in the regulation of Ia gene expression throughout B-cell development and in response to inducers. As the immune response is a phenomenon involving interacting cells and multiple coordinately expressed gene products, we examined Ia expression and function in the whole organism by injecting Ia genes into mouse single-cell embryo pronuclei and identifying the resultant transgenic mice. Recent experiments injecting immunoglobulin, elastase and transferrin genes and some genes under the control of the metallothionein promoter have documented expression of these genes in a tissue-specific or at least a tissue-preferential fashion^{6–14}.

Here we demonstrate the use of transgenic mice in studying Ia expression and function. Mice of the $H-2^s$ or $H-2^b$ haplotypes do not express the E_α gene because of a deletion that removes its promoter region and first exon^{15,16}; consequently they cannot produce antibody in response to

E-restricted antigens such as the synthetic polymer poly(Glu⁵⁷Lys³⁸Phe⁵) (GLØ). We have injected the E_α^k gene into $H-2^{(b \times s)}$ hybrid mouse embryos. The resulting transgenic mice transcribe the E_α^k gene in a tissue-specific manner, express heightened levels of E_α in response to IFN- γ and assemble an $E_\alpha^k/E_\beta^{b \text{ or } s}$ complex at the cell surface that can be recognized by monoclonal antibodies. Finally, as evidence of the overall efficacy of the injected E_α^k gene, the transgenic mice can produce antibody in response to GLØ.

E_α^k gene incorporation

The E_α^k gene from strain A/J mice has previously been cloned into λ Charon 4 (ref. 17). A partial restriction enzyme map of this clone is presented in Fig. 1a, which also shows the position of a 600-base pair (bp) deletion at the 5' end of the E_α gene from $H-2^b$ and $H-2^s$ mice. This deletion precludes E_α gene transcription in these strains because it includes the promoter region and the first exon¹⁶.

We injected the cloned E_α^k gene into the male pronucleus of F_2 hybrids derived from C57BL/6 $\times$ SJL hybrid crosses. (Note that SJL mice are $H-2^s$ and C57BL/6 are $H-2^b$.) This strain combination was chosen because it produces vigorous embryos⁷, is easy to manipulate and the progeny, $H-2^b$ and/or $H-2^s$, would not transcribe the endogenous E_α gene. In one experiment we injected the entire λ clone; 21 offspring developed from 88 surviving embryos, one of which (no. 23) had incorporated E_α^k . In a second experiment, we injected the purified 8.2-kilobase (kb) *Bgl*I fragment which contains all the E_α exons plus 2 kb of DNA 5' to the cap site and 1.4 kb 3' to the poly(A) addition site (see Fig. 1a). Of 273 surviving embryos, 16 were born live, including one (no. 16) that had incorporated E_α^k .

Figure 1b shows a Southern blot of tail DNA digested with *Bam*HI/*Eco*RV and hybridized with a small fragment (probe 1) entirely contained within the E_α deletion of $H-2^b$ and $H-2^s$ mice. Only the E_α^k gene should generate a signal with this probe (see Fig. 1a). Mice Ea16 and Ea23 exhibit a *Bam*HI/*Eco*RV fragment indistinguishable in size from that of A/J mice which contain two E_α^k genes per diploid genome. In contrast, Ea7 did not incorporate the injected DNA and thus provides a useful negative control.

The blot in Fig. 1c shows *Hind*III/*Eco*RV digests hybridized with probe 2, identified in Fig. 1a. Lane 1 of Fig. 1c shows the pattern expected from an E_α^k gene: a large band of ~6.6 kb containing all the E_α exons and a small band of ~1.8 kb containing sequences 5' to the coding region. The $E_\alpha^{b \text{ or } s}$ patterns are seen in the next two lanes. The large band is smaller than for E_α^k because of the 600-bp deletion; the small band is larger than for E_α^k because of a polymorphic *Hind*III site (D.M. and C.B., unpublished observations). Mouse 23 exhibits all the bands characteristic of the E_α^k and $E_\alpha^{b \text{ or } s}$ genes, whereas mouse 16 has the large bands diagnostic of E_α^k and $E_\alpha^{b \text{ or } s}$ and the small band of only the latter. The small E_α^k band was not expected because the 5' border of the *Bgl*I fragment injected into mouse 16 is contained in the small *Hind*III/*Eco*RV fragment (see Fig. 1a).

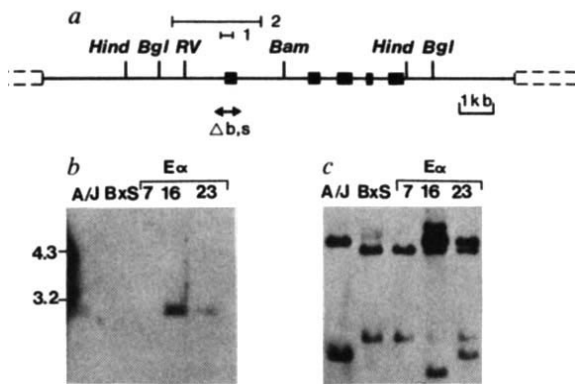


Fig. 1 Integration of the E_α^k gene into the genome of transgenic mice. *a*, Map of the E_α^k gene in the λ clone $\lambda A/J$ p34.13 (ref. 17). Thick lines, exons; thin lines, introns and flanking sequences; dashed boxes, vector arms. The deletion of the $H-2^b$ and $H-2^s$ haplotypes in E_α is shown as a double-headed arrow. Brackets indicate the two probes used. Probe 1, a 320-bp *AccI*/*SacI* fragment (from position -220 to +100 with respect to the major site of initiation of transcription) that does not hybridize to DNA from $H-2^b$ and $H-2^s$ haplotype mice. Probe 2, a 2-kb *SalI* fragment spanning ~1 kb on either side of the b/s deletion. Only the relevant restriction enzyme sites are shown. *Bam*, *Bam*HI; *Bgl*, *Bgl*II; *Hind*, *Hind*III; *RV*, *Eco*RV. *b*, *c*, Southern blot analyses (described in ref. 17) of tail DNA from several mice: A/J, C57BL/6 $\times$ SJL F₁ (B $\times$ S), and mice derived from injected eggs (E α 7, 16 and 23). In *b*, DNA was digested with *Bam*HI and *Eco*RV, and the blot hybridized to probe 1. *c*, The pattern generated with *Hind*III and *Eco*RV was revealed by probe 2. The band of high relative molecular mass in E α 16 results from incomplete digestion.

It is replaced by a slightly smaller band, whose size is consistent with the notion that many of the copies of E_α^k in mouse 16 are organized as tandem repeats. This blot demonstrates first, that the injected genes are not grossly rearranged in either mouse 23 or mouse 16, as shown by the intact 6.6-kb band containing all the exons. Second, by comparing within the same slot, the band intensities of fragments derived from the injected E_α^k genes with those from the endogenous $E_\alpha^{b,s}$ genes, we can accurately quantitate E_α^k copy number. Thus, mice 23 and 16 contain, respectively, ~4 and 12 copies of the intact E_α^k gene per diploid genome in tail tissue. Similar results were obtained with liver DNA.

To ascertain whether these genes were incorporated in the germ line, we mated mice 23 and 16 with mice from strain C57BL/6 ($H-2^b$). Mouse 23 produced nine young, four of which had integrated the E_α^k gene, whereas 7 of the 16 offspring from mouse 16 had the E_α^k gene. These values suggest that, for each transgenic mouse, all copies of the injected gene had integrated into the same chromosome. Southern blots of *Bam*HI digests confirmed that the offspring from mouse 16 had inherited the same number of copies, and that this number was the same as for their parent (data not shown).

E_α^k expression

The E complex is absent from most murine cells, but is present in high concentrations on B cells and activated macrophages⁴. Thus, we chose spleen tissue for the initial analysis of transcription from the injected E_α^k gene. A Northern blot of total spleen RNA hybridized with an E_α complementary DNA probe (Fig. 2a) shows that transgenic mice E α 16 and E α 23 express the same 1.25-kb E_α transcripts as control mice of haplotype $H-2^{(b \times a)}$. It should be recalled that neither the E_α^b nor E_α^s genes are transcribed, so that all the visualized RNA derives from the E_α^k gene.

To confirm the accuracy of transcription, we mapped initiation sites on the E_α^k gene in the spleen. RNA was hybridized with a single-stranded DNA probe labelled at the 5' end. If initiation were correct, treatment of the DNA-RNA hybrids with S₁ nuclease should result in a pair of protected fragments ~100 bp

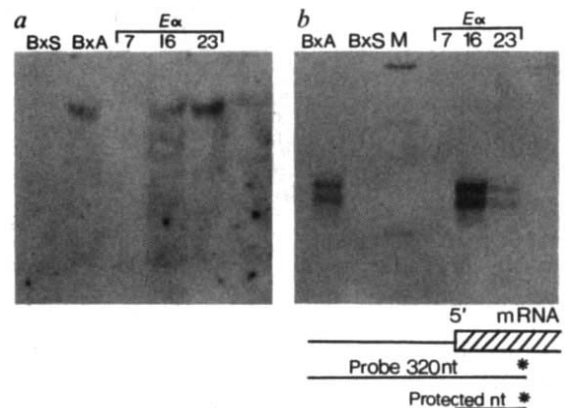


Fig. 2 Accurate expression of the integrated E_α^k gene. Total RNA was prepared from spleens of different control or transgenic mice and analysed by Northern blotting (*a*) or S₁ nuclease mapping (*b*). The nomenclature for the mouse strains is described in Fig. 1 legend, except for B $\times$ A which is a C57BL/6 $\times$ A/J F₁ hybrid. The filter in *a* was probed with an E_α cDNA clone¹⁷. In *b*, the probe was a 320-nucleotide (nt) single-stranded DNA fragment (*AccI*/*SacI*; see Fig. 1 legend). Hybridization with E_α mRNA protects a set of fragments ~100 nucleotides long, which was revealed by electrophoresis on a urea-acrylamide gel followed by autoradiography. M, relative molecular mass markers with sizes of 147, 122, 110, 90 and 76 nucleotides. Techniques and conditions have been described elsewhere¹⁷.

long, as is seen in Fig. 2b for a B $\times$ A mouse. The complexity of the pattern results from multiple initiation sites (see discussion in ref. 17). In spleen RNA from mice 16 and 23, initiation on the E_α^k gene was accurate. Mouse 23 seems to produce fewer transcripts than the B $\times$ A hybrid, whereas mouse 16 makes more. These differences are more apparent than in the Northern blot, probably because the S₁ assay is less sensitive to partial RNA degradation.

The specificity of transcription was evaluated by S₁ analysis of RNA from several other tissues (Fig. 3). In the B $\times$ A control, high levels of E_α messenger RNA were observed in spleen and lung tissue, lower levels were detected in the kidney and almost no E_α mRNA could be measured in the liver, heart and brain. The results with lung and kidney are somewhat surprising in view of the contention that Ia genes are expressed mainly on B cells and macrophages. However, Northern blots of lung and kidney RNA hybridized with an immunoglobulin κ -chain probe indicated that these tissues are contaminated by lymph nodes, macrophages and/or significant blood infiltration (data not shown; ref. 14). Transgenic mice E α 16 and E α 23 exhibit the same pattern of tissue distribution: high levels in spleen and lung and low or undetectable levels in kidney, liver, heart and brain (Fig. 3). This S₁ assay may not detect aberrant transcripts initiating in regions of the E_α gene not covered by the rather short probe. To exclude this possibility, we also ran Northern blots on RNA from the different tissues and did not detect anomalous bands, confirming the S₁ mapping results (data not shown).

Another method of assessing tissue-specific expression of the injected gene is to compare the amount of A_α and E_α^k transcripts in various tissues. A_α has a high level of sequence homology with, an identical tissue distribution to, and a functional role analogous to the E_α chain. In S₁ experiments similar to that illustrated in Fig. 3, RNA was hybridized simultaneously with E_α and A_α probes (see Fig. 3 legend). Clearly, the bands attributed to E_α^k mRNA (100 bp) have the same pattern of distribution as those resulting from A_α transcripts (128 and 260 bp). We concluded that the injected E_α^k gene is transcribed coincidentally with the endogenous A_α gene in the six tissues studied. In other experiments, we have confirmed this tissue-specificity of expression in one offspring from each of mice 16 and 23.

Evidence for the tissue-specificity of transcription is

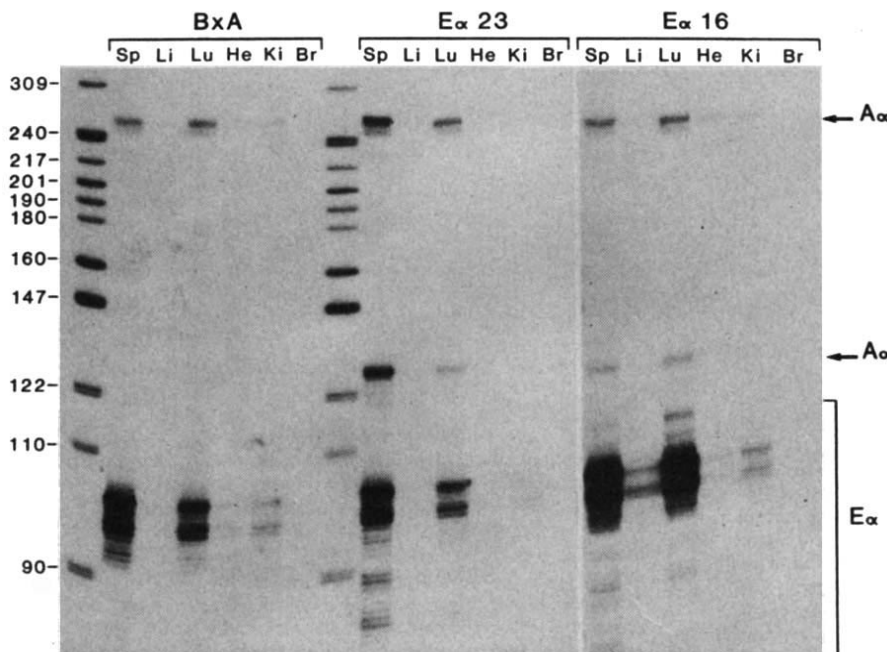


Fig. 3 Tissue-specific expression of E_{α}^k in transgenic mice. E_{α} and A_{α} mRNA was analysed in various tissues from three mice by S_1 nuclease mapping of total RNA. For E_{α} , the probe was the same as in Fig. 2. For A_{α} , the probe was an 800-bp *Hind*III fragment from an A_{α}^b cDNA clone, which was labelled at the *Hind*III site at position 262 and extended through the first domain, the leader and the plasmid sequences. A_{α}^b mRNA protects a 262-nucleotide fragment. The second A_{α} band on the gel results from an internal S_1 nuclease cut when the probe is hybridized to A_{α}^s mRNA; allelically polymorphic residues of the s haplotype²¹ locally destabilize the hybrid enough for partial S_1 nuclease cleavage under our conditions. The tissues investigated were spleen (Sp), liver (Li), lung (Lu), heart (He), kidney (Ki) and brain (Br), from a C57BL/6 $\times$ A/J F_1 hybrid (B $\times$ A), and from E α 23 and E α 16.

thus twofold: first, E_{α}^k expression in transgenic mice closely parallels that in the B $\times$ A control; second, transcription of the E_{α}^k and the endogenous A_{α} genes follows a very similar pattern. However, because some of the tissues exhibit a high background resulting from blood infiltration, macrophages or lymph nodes, we cannot eliminate the possibility that the transgenic mice exhibit low levels of aberrant transcription in these tissues. This might be the case in mouse 16 liver, where a low level of E_{α} mRNA (a few per cent of that seen in the spleen) is observed.

As a measure of cell-type specificity of expression, we compared the distribution of the A and E complexes on spleen cells by labelling them with fluorescent anti-E and anti-A monoclonal antibodies and visualizing them by two-colour cytofluorometric analysis. Expression of the A complex on the cell surface depends on transcription of the endogenous A_{β} and A_{α} genes; expression of E requires transcription of the endogenous E_{β} and the injected E_{α} genes. Thus, we could not ask whether E_{α}^k was transcribed in cells in which the other Ia chains did not occur because we would not detect it on the cell surface (resulting from the absence of E_{β}). However, we could determine whether cells expressing the A complex (and thus also transcribing the endogenous E_{β}) all exhibit an E complex on the cell surface. Figure 4 shows that the spleen cells of one of the progeny of mouse 16 consist almost entirely of two populations: $A^{-}E^{-}$ and $A^{+}E^{+}$. As expected, there is no $A^{+}E^{-}$ component. The fact that there is no significant $A^{+}E^{-}$ population indicates that essentially all cells that synthesize the endogenous A_{α} , A_{β} and E_{β} chains also transcribe E_{α}^k . A few cells trail into the $A^{+}E^{\text{null}}$ area, but this minor component can also be seen in spleens from B $\times$ A controls.

E_{α}^k gene induction

Expression of all four Ia antigens is stimulated greatly in macrophages by treatment with supernatants from activated T cells or with purified IFN- γ . This induction was first described for cell-surface Ia molecules (see ref. 18 for review), but more recently has been documented for stable mRNA transcripts^{17,19}. To determine whether the injected E_{α}^k gene was similarly sensitive to IFN- γ induction, we prepared peritoneal macrophages from E α 16 and E α 23 third-generation mice, cultured them in the presence or absence of recombinant IFN- γ for 60 h and quantitated A_{α} and E_{α} mRNA levels by S_1 mapping.

Figure 5 clearly shows an induction of A_{α} and E_{α}^k mRNA in the IFN- γ -treated macrophages from both transgenic mice. The levels of Ia induction are similar to those that we have observed for E_{α}^k in repeated experiments on other mice.

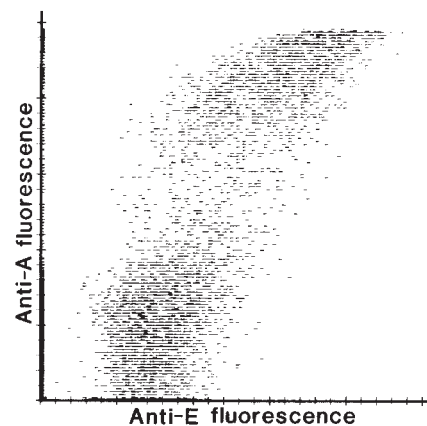


Fig. 4 I-A-positive spleen cells are also I-E-positive in the E α 16 transgenic line. Twenty hours after concanavalin A activation, spleen cells were stained with fluorochrome-labelled monoclonal antibodies specific for the A or the E complex. For each cell, red and green fluorescence was detected by flow cytofluorometry. On this two-parameter representation, single cells appear as thin dots; thicker dots indicate several cells in the same channel.

Methods. Antibodies were purified from ascites fluid by affinity chromatography on protein A-Sepharose (Pharmacia). The anti-I-A antibody BP107 (ref. 22) was directly conjugated to fluorescein isothiocyanate (FITC) as described elsewhere²³. Antibody 14.4.4, specific for the E complex, was conjugated to phycoerythrin (PE; Molecular Probes, Ore) according to Oi *et al.*²⁴. After purification on Lympholyte M (Cedarlane), lymphocytes were stained in one step in FWS (phosphate-buffered saline supplemented with 3% calf serum, 30 mM HEPES and 0.1% sodium azide). Incubation was for 30 min on ice with saturating amounts of antibody. After two washes in FWS, the cells were examined on an ATC 3000 flow cytometer (ODAM, Wisselbourg) equipped with an Innova-90 argon laser set on the 488-nm line for excitation of both fluorochromes. Signals from lymphocytes were selected by electronic gating according to size (impedance measurement) and laser scatter. Emission from the fluorochromes was separated by a 555-nm SWP017 dichroic mirror (Melles Griot, Irvine) and further elimination of signal cross-contamination was accomplished with a 550-nm SWP015 short-pass filter in front of the FITC detector and a 575-nm 5151 long-pass filter (Oriel, Stratford) in front of the PE detector. A small degree of electronic compensation was also applied before logarithmic amplification. Average channel numbers on the red and green scales were: $A^{-}E^{-}$, 73 and 52; $A^{+}E^{+}$, 160 and 210; unstained cells (or cells stained with irrelevant antibodies) were centred around channels 71 and 45 in the red and green scales, respectively.

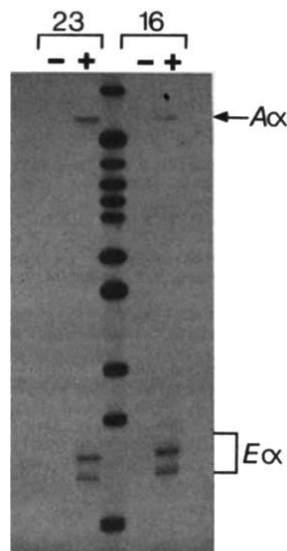


Fig. 5 Injected E_{α}^k gene response to IFN- γ . Peritoneal macrophages were elicited with thioglycolate (3 days)²⁵ and subsequently cultured for 60 h in the presence (+) or absence (-) of 100 U ml⁻¹ recombinant IFN- γ . RNA was prepared and analysed by S₁ nuclease mapping as described in Fig. 3 legend. The mice were from the F₂ of a backcross to C57BL/6.

Functional E_{α}^k gene product

The ability of an animal to make antibody in response to the synthetic antigen GL $\emptyset$ is dependent on expression of the E complex. The mice that served as hosts for the injected E_{α}^k gene do not respond to GL $\emptyset$ because they are the progeny of (H-2^b × H-2^s) hybrid crosses and thus do not express any E_{α} . However, if the injected E_{α}^k gene makes a functional product, the animal should be able to mount a response to GL $\emptyset$. To test this possibility, we injected some of the offspring of mice 16 and 23 with GL $\emptyset$ according to the protocol described in Fig. 6 and evaluated their secondary response by quantitating serum levels of anti-GL $\emptyset$ antibody. Note that these offspring derive from a C57BL/6 × transgenic mating, so only about half carry the injected E_{α}^k gene. Ovalbumin was injected at the same time to serve as a positive control; as it is an A-restricted antigen, b- and s-haplotype mice can respond.

Figure 6 shows a perfect correlation between the presence of the E_{α}^k gene and the ability to make antibody in response to GL $\emptyset$. The titres and mRNA levels for the offspring of mouse 16 are generally higher than those of mouse 23 (Figs 2b, 3). There is some variability in the response for different mice from the same transgenic parent, but we have also seen variable antibody titres for the B × A controls, and thus we do not lend much significance to the range of responses. The lower values cannot be attributed to the H-2 haplotypes (see Fig. 6, beneath the histogram), to a difference in the number of E_{α}^k genes that have been inherited (data not shown) or to general immunodeficiency (because of a healthy ovalbumin response).

Conclusion

We have injected a cloned E_{α}^k gene into embryos of mice incapable of expressing their endogenous E_{α} genes. The cloned DNA was incorporated into the germ line and inherited in a normal mendelian fashion. Expression of the E_{α}^k gene is tissue-specific, similar to transcription of the endogenous homologue, A_{α} . In peritoneal macrophages, E_{α} mRNA synthesis is responsive to induction by IFN- γ . The injected gene is transcribed accurately and specifically in several cell types, and the animals can mount a normal immune response.

Our results show that tissue-specific expression and inducibility by IFN- γ of the E_{α} gene is conferred by sequences within or relatively close to the structural gene: mouse 16 was injected with an 8.2-kb fragment containing only 2 kb upstream from the cap site and 1.4 kb downstream to the poly(A) addition site.

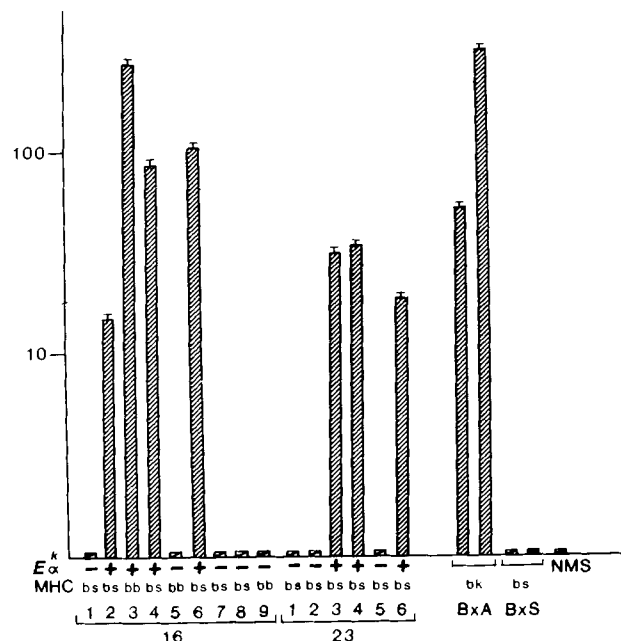


Fig. 6 E_{α}^k -positive transgenic mice mount an effective immune response to the synthetic antigen GL $\emptyset$. Fifteen mice, derived from crossing $E_{\alpha}16$ or $E_{\alpha}23$ with C57BL/6, were challenged with the terpolymer (Glu-Lys-Phe), and specific antibody titres were measured in a sandwich enzyme-linked immunosorbent assay. Bars represent the values (in arbitrary units) obtained for the different offspring and for control B6 × A/J or B6 × SJL F₁ hybrids. MHC haplotypes are shown below the E_{α}^k -status, and were determined for the transgenics by Southern blotting of genomic DNA cut by *Pst*I and probed with an A_{α} cDNA fragment. *Pst*I generates polymorphic A_{α} bands which enable the H-2^b and H-2^s haplotypes to be distinguished. NMS, normal mouse serum.

Methods. Mice were taken at 2 months old and were injected intraperitoneally at days 1 (in complete Freund's adjuvant) and 21 (in saline), each time with 25 μ g GL $\emptyset$ and 25 μ g chicken ovalbumin (OVA). Sera were collected at day 28 and tested for specific antibodies by incubating dilutions in microtitre plates previously coated with GL $\emptyset$ or OVA. After washing off unbound antibodies, we added to the wells an alkaline phosphate-conjugated goat anti-mouse immunoglobulin antibody (Cappel); this was followed after 2 h by extensive washing. Bound enzyme was measured by the rate of conversion of *p*-nitrophenyl phosphate, with several absorbance readings taken on a Titertech Multiscan. Titres were calculated from the linear portions of the curves obtained, and are expressed in arbitrary units by comparison with a laboratory standard serum.

Therefore, in future experiments it will not be necessary to inject large cosmid clones that may contain adjacent genes.

By the injection of a single gene we have endowed a genetically deficient host strain with a new immune response. A successful antibody response implies the involvement of the E complex in various cell/cell interactions, for example, APC-T cell, T-B cell. In addition, T cells must be 'educated' to the E complex in the thymus. That the E_{α}^k gene product seems to function correctly in all the necessary immunological contexts allows the study of Ia function in the organism. The importance that mutant and recombinant mice have already had in studies elucidating the intricacies of the immune response cannot be underestimated. The generation of such strains is, however, limited to fortuitous mutations or crossovers, and unknown genetic segments can be carried along during the construction of lines (see, for example, ref. 20). It should now be much easier to create mouse strains which differ by only a single gene.

We thank R. Brinster, K. Illmensee, R. Jaenisch, D. Martin and V. Williams II for advice, Y. Quesneau for the cytofluorometric analysis, C. Irrmann for reimplanting the zygotes, and P. Gerber and P. Hickel for technical assistance. We also thank H. McDevitt, J. Monaco, G. Adolph, D. Landais, M. Ayane, G. Garcin and S. Candeias for gifts of materials and A. Staubli for tissue sections. Recombinant IFN- γ (produced by Genentech Inc.) was supplied by Boehringer Ingelheim. E.

Taubert, S. Phayboune, C. Marfing, and F. Mackay-Landau were responsible for building and maintaining the animal care facility. J. Kaibach, B. Boulay and C. Werlé assisted in the preparation of the manuscript. Finally, we thank P. Chambon for his continued support and interest and for having the

foresight to develop an injection facility. This work was supported by funds from CNRS (ATP 6182), Institut National de la Santé et la Recherche Médicale, Association de la Recherche sur le Cancer, Fondation de la Recherche Médicale Française and Fondation Simone et Cino del Duca.

Received 22 March; accepted 10 May 1985.

1. Klein, J. in *Immunology, The Science of Self-Nonself Discrimination*, 270-309 (Wiley, New York, 1982).
2. McDevitt, H. O. *J. Immunogenet.* **8**, 287-295 (1981).
3. Schwartz, R. H. in *Antigens* Vol. 1 (eds Ferrone, S. & David, C. S.) 161-218 (CRC, Boca Raton, Florida, 1982).
4. Frelinger, J. A. in *Antigens* Vol. 1 (eds Ferrone, S. & David, C. S.) 37-54 (CRC, Boca Raton, Florida, 1982).
5. Mingle-Gaw, L. & McDevitt, H. O. *A. Rev. Immun.* **3**, 369-398 (1985).
6. Brinster, R. L. *et al. Cell* **27**, 223-231 (1981).
7. Palmiter, R. D., Chen, H. Y. & Brinster, R. L. *Nature* **300**, 611-615 (1982).
8. McKnight, G. S., Hammer, R. E., Kuenzel, E. A. & Brinster, R. L. *Cell* **34**, 335-341 (1983).
9. Brinster, R. L. *et al. Nature* **306**, 332-336 (1983).
10. Grosschedl, R., Weaver, D., Baltimore, D. & Constantini, F. *Cell* **38**, 647-658 (1984).
11. Storb, U., O'Brien, R. L., McMullen, M. D., Gollahon, K. A. & Brinster, R. L. *Nature* **310**, 238-241 (1984).

12. Swift, G. H., Hammer, R. E., MacDonald, R. J. & Brinster, R. L. *Cell* **38**, 639-646 (1984).
13. Ornitz, D. M. *et al. Nature* **313**, 600-602 (1985).
14. Chada, K. *et al. Nature* **314**, 377-380 (1985).
15. Mathis, D. J., Benoist, C., Williams, V. E., Kanter, M. & McDevitt, H. O. *Proc. natn. Acad. Sci. U.S.A.* **80**, 273-277 (1983).
16. Dembic, Z. *et al. EMBO J.* **4**, 127-131 (1985).
17. Mathis, D. J., Benoist, C. O., Williams, V. E., Kanter, M. R. & McDevitt, H. O. *Cell* **32**, 745-754 (1983).
18. Unanue, E. A. *Rev. Immun.* **2**, 295-428 (1984).
19. Nakamura, M., Manser, T., Pearson, G. D. N., Daley, M. J. & Geffer, M. L. *Nature* **307**, 381-382 (1984).
20. Hayes, C. E. *et al. Science* **223**, 559-563 (1984).
21. Landais, D., Matthes, H., Benoist, C. & Mathis, D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2930-2934 (1985).
22. Ozato, K., Mayen, N. & Sachs, D. J. *Immun.* **124**, 533-541 (1980).
23. Symington, F. & Sprent, J. *Immunogenetics* **14**, 53-61 (1981).
24. Oi, U., Glaser, A. & Stryer, L. *J. Cell Biol.* **93**, 981-988 (1982).
25. Gallily, R. & Feldman, H. *Immunology* **12**, 197-205 (1967).

LETTERS TO NATURE

On the nature of the companion to Van Biesbroeck 8

L. A. Nelson, S. A. Rappaport & P. C. Joss

Center for Theoretical Physics and Department of Physics,
Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139, USA

We present here the results of the first numerical evolutionary calculations for very low-mass stars (masses in the range of $0.01-0.1 M_{\odot}$) with ages up to the age of the Galaxy. Our calculations shed new light on the nature of the recently discovered object VB 8B (ref. 1), the companion to Van Biesbroeck's star no. 8 (VB 8; ref. 2), and support the identification¹ of this object as a low-mass star in which hydrogen burning was never able to establish thermal equilibrium. For plausible ages of this binary stellar system ($\geq 10^9$ yr), we estimate the mass, M , of VB 8B to be in the range $0.04-0.08 M_{\odot}$ (see also ref. 1 and refs therein). The object is supported primarily by degenerate electron pressure, and we find that its present values of central density and temperature are $\rho_c \approx 500-2,000 \text{ g cm}^{-3}$ and $T_c \approx 1.0-1.5 \times 10^6 \text{ K}$, respectively. The star VB 8 itself³⁻⁵ may also be a low-mass star which marginally failed to establish main-sequence hydrogen burning. Thus, the mass ratio for this binary system is probably not greatly different from unity, and we predict an orbital period of ≥ 40 yr. Our evolutionary tracks are also applicable to the interpretation of observations of other very low-luminosity dwarfs (see refs 3, 5-7). The study of such objects may be directly relevant to the development of an understanding of the 'dark matter' that has been inferred, from dynamical considerations, to be present in our Galaxy and many extragalactic systems (see, for example, refs 8, 9).

Several stars have been found to have either a mass close to or below that of the end of the hydrogen-burning main sequence (for example, Luyten 726-8A and B; Ross 614B; refs 6, 7, 10-12), or an effective temperature below that expected for even the lowest mass main-sequence star (for example, LP 271-25 = LHS 2924; ref. 5). Some of these stars are in binary systems whereas others have no detected companion. Such stars may be of sufficiently low mass ($\leq 0.08 M_{\odot}$ for a hydrogen-rich star^{13,14}) that they never achieve thermal equilibrium through hydrogen burning, and are candidates for 'brown dwarfs'. This is something of a misnomer, in that some stars with masses below the minimum main-sequence mass can attain effective temperatures quite comparable with those of ordinary late M dwarfs for a substantial interval of time ($\geq 3 \times 10^8$ yr) during their contraction phase. The existence of very low-mass ($\leq 0.01 M_{\odot}$), unseen companions to nearby stars has also been inferred by astrometric

techniques^{15,16}, but the reliability of these detections remains uncertain (see ref. 17).

In addition, the presence of low-mass ($M \leq 0.1 M_{\odot}$) stars has been inferred in several interacting binary systems, including two binary X-ray sources and several cataclysmic variables with orbital periods near the minimum in their orbital period distribution (see refs 14, 18, 19 and refs therein). In all cases, however, these binary systems are studied in X-radiation and/or visible light produced by the accretion of matter from the hypothesized low-mass secondary to a collapsed star (degenerate dwarf or neutron star). The low-mass star itself is not directly detected. Furthermore, in these systems, the secondary stars are thought to have attained their present low-mass state by transferring a large amount of their original mass to the collapsed star (or by loss of mass from the binary altogether). Thus, most of these low-mass interacting stars are probably not primordial: that is, they have evolved from more massive objects.

To get a better understanding of the object VB 8B, which is reported to have a radius, R , of $\sim 0.09 R_{\odot}$ and an effective temperature, T_e , of $\sim 1,360 \text{ K}$ (ref. 1), we have carried out evolutionary calculations for stars of mass $\leq 0.1 M_{\odot}$. We utilized a simplified stellar evolution code developed by Rappaport *et al.*¹⁹ and modified by Rappaport and Joss¹⁴ to carry out the present calculations. The basic simplifying features of the code are in (1) the treatment of the stellar structure, which is taken to be completely convective and is represented by an $n = 3/2$ polytrope, and (2) the matching of entropy in the stellar interior directly to that at the photosphere. The former approximation has been shown to be adequate^{6,13,18-20} for a low-mass star during its cooling from very early evolutionary stages all the way to a totally degenerate configuration. The assumption of matched entropies has been utilized successfully in the past¹⁴, and we have carried out further checks in the present work (see below). The code we utilize, however, cannot handle cold objects of mass $\leq 0.005 M_{\odot}$. For the present study, we have slightly modified the code to include the burning^{7,21} of a small quantity (5×10^{-5} by mass; see ref. 22) of primordial deuterium, and an approximate expression which incorporates the Coulomb corrections to the equation of state^{23,24}. This latter expression was adjusted so that the stellar radii we obtain for totally degenerate configurations almost exactly match the results of more detailed structure calculations by Zepolsky and Salpeter²⁵. Unlike some previous calculations of cooling curves for very low-mass stars (see 6, 7, 26 and refs therein), our evolutionary runs are continued to very low effective temperatures and the behaviour of the parameters is presented explicitly as functions of time. Stevenson²⁷ has carried out semi-analytical calculations of the cooling of degenerate dwarfs to very low temperatures, but, unlike the present study, his calculations did not follow the evolution of the stellar radius.

We considered stellar masses in the range of 0.01 – $0.10 M_{\odot}$, in increments of $0.01 M_{\odot}$. We note, that $0.01 M_{\odot}$ is an especially plausible operational definition of the dividing line between 'stars' and 'planets', because this is about the minimum mass for which the pressure support in the final equilibrium configuration is contributed principally by either thermal gas pressure or degenerate electron pressure, as opposed to atomic and molecular forces. Moreover, $0.01 M_{\odot}$ is about the minimum mass for which quasi-equilibrium thermonuclear fusion (in the form of deuterium burning) can ever be established (see below).

We tested two chemical compositions, one appropriate to Population I stars $[(X, Y, Z) = (0.70, 0.28, 0.02)]$ and the other appropriate to Population II $[(X, Y, Z) = (0.719, 0.28, 0.001)]$. We used three sets of radiative opacities; one^{19,28} simulated the effects of grain formation in the stellar atmosphere at low temperatures ($T_e \leq 2,200$ K; hereafter, opacity AG), while the other two, which ignored such effects, were adapted from the opacity tables derived by Alexander^{19,28} and the fitting formulae, due to Webbink (ref. 14 and personal communication), based on the Cox–Stewart²⁹ and Cox–Tabor³⁰ opacities (hereafter, opacities A and W, respectively). The effect of possible grain formation on the radiative opacities at low temperatures was approximated by setting the numerical value of the Alexander^{19,28} opacity coefficient equal to a constant for $T_e \leq 2,200$ K (see also ref. 18).

All initial stellar models had $R \geq 1 R_{\odot}$, which is at least 10 times greater than the radius of any plausible model of VB 8B, and were assumed to have negligible rotational angular momentum. If, however, the angular momentum were not significantly dissipated by the time the contracting protostar had reached a radius of $\sim 1 R_{\odot}$, then the early portions of the evolutionary cooling tracks could be quite different from those we present. On the other hand, it is quite plausible that some form of braking torque, possibly associated with a magnetic stellar wind (see ref. 31), will effectively despin the star, so that for stellar ages $\geq 10^7$ yr our evolutionary calculations should be valid. (For a recent review of some of the processes involved in star formation, see ref. 32 and refs therein.)

Figure 1 shows cooling curves for stellar models of several different masses; Fig. 1a and b show the evolution of the effective temperature and stellar radius, respectively. All models in Fig. 1 were generated with a Population I chemical composition and the AG radiative opacities. Note that the contraction of each stellar model is temporarily halted for times between $\sim 10^6$ and $\sim 10^7$ yr due to deuterium burning (see also ref. 7). After deuterium exhaustion, contraction resumes towards the final equilibrium state. For $M \geq 0.086 M_{\odot}$, the star achieves a state of thermal equilibrium (that is, arrives on the hydrogen-burning main sequence) within $\sim 10^9$ yr and has an effective temperature far greater than that measured for VB 8B [$1,360 \pm 200$ K (ref. 1)]. For lower mass stars, thermal equilibrium is never reached (except temporarily on the deuterium-burning main sequence) and the stars cool indefinitely. However, for a star whose mass is only slightly lower than the minimum mass for the hydrogen-burning main sequence, cooling times required to achieve $T_e \leq 2,000$ K may exceed the age of the Universe. We note that for $T_e \leq 2,000$ K the atmospheric radiative opacities become highly uncertain; however, we have found that the cooling timescales are rather insensitive to changes in the assumed opacity law.

We have repeated the evolutionary calculations presented in Fig. 1 for the other radiative opacities discussed above and for stars with a Population II chemical composition. The results are in qualitative agreement with those shown in Fig. 1. The implementation of opacity A has the effect of initially decreasing the cooling rate and then increasing the rate at later times. We found that the greatest overall change in the cooling timescale, relative to that obtained with the AG opacities, was about a factor of 3 and occurred for the $M = 0.01 M_{\odot}$ model. More typically, for other masses, the cooling times changed by a factor of < 2 . When opacity W is used, the change in the behaviour of the cooling rate is similar to that found for opacity A, and the effective temperatures are systematically raised by ~ 10 – 15% during the initial contraction phase. As a further test of the

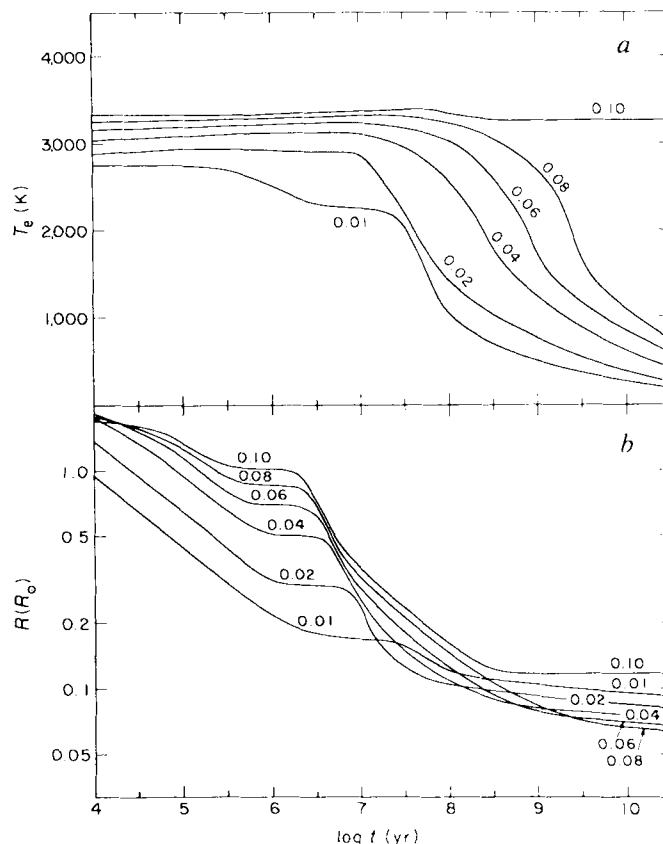


Fig. 1 Evolutionary tracks for very low-mass stars with masses in the range of 0.01 – $0.10 M_{\odot}$. The effective temperature, T_e , and stellar radius, R , are shown as functions of elapsed time, t , in a and b, respectively. Each track is labelled with the corresponding stellar mass, in units of $M_{\odot}$. The plateau regions evident in the plots of radius versus time (near $t \approx 10^6$ – 10^7 yr) are due to the burning of a small quantity (5×10^{-5} by mass) of primordial deuterium. Stars with masses in excess of $\sim 0.086 M_{\odot}$ ultimately achieve thermal equilibrium through hydrogen burning (that is, they attain hydrogen-burning main-sequence status). Stars of lower mass cool indefinitely towards a completely degenerate configuration. For $T_e \leq 1,000$ K the displayed cooling curves become increasingly uncertain. Our values of T_e for $t \geq 3 \times 10^9$ yr are in good agreement with those calculated by Stevenson²⁷.

sensitivity of our results to the values of the radiative opacities at low temperatures, we have calculated several series of cooling curves where we adopted various (constant) values for the opacity coefficient at $T_e < 2,200$ K. We find that for changes in this constant opacity of up to two orders of magnitude, the cooling times change by only a factor of ≤ 2 . We caution, however, that extremely large changes in the opacity law might alter the stellar structure and surface-boundary conditions sufficiently to invalidate the simplifying assumptions that we have utilized in our code. Finally, in a similar series of tests, we have verified the insensitivity of our results to the assumption that the entropy in the stellar interior is equal to that at the photosphere.

We note that additional complications in the physics of the stellar interior, such as crystallization into a Coulomb lattice (see ref. 33) within some layers of the star, may occur at the lowest effective temperatures we consider. We have checked, in a self-consistent manner, that such effects should not affect our basic results obtained with the AG opacities for ages of $\leq 10^{10}$ yr. The effects of crystallization become important earlier during the evolution when the A or W opacities are used than when the AG opacities are employed.

If we assume that VB 8B is not exceptionally young (that its age is $\geq 10^9$ yr; ref. 1), then we can conclude from Fig. 1 that its mass must exceed $\sim 0.04 M_{\odot}$ in order for it not to have cooled below its current effective temperature. (However, observational

selection might favour the discovery of such objects when they are young and thus relatively bright; if this is the case, somewhat smaller values for the age of VB 8B might be allowed.) The effective temperature of a star with a mass $\geq 0.086 M_{\odot}$ is always much higher than that observed for VB 8B. For stars with masses in the range of ~ 0.04 to $\sim 0.08 M_{\odot}$, the radii and effective temperatures will be consistent with those observed for VB 8B during a substantial interval of time. Table 1 lists some of the physical properties of our stellar models in this mass range that have radii and effective temperatures similar to those of VB 8B.

It seems quite reasonable that VB 8 itself is also an extremely low-mass star which has a mass very close to that of a star at the end of the hydrogen-burning main sequence. According to Probst and Liebert⁵, its effective temperature is $\sim 2,400$ K. [We note, however, that temperature determinations in this range are highly uncertain, and the effective temperature of VB 8 could be somewhat higher (see ref. 1).] Table 1 also lists the properties of our stellar models with temperatures close to that observed for VB 8. A star with a mass of $\sim 0.08 M_{\odot}$ and an age of $\sim 2 \times 10^9$ yr would have properties consistent with those observed for VB 8 (refs 2–5). The absolute bolometric magnitude of such a star is 14.2. This value, together with the observed visual magnitude of ~ 16.9 (refs 2, 3) and the measured distance of 6.5 pc (ref. 2), implies a bolometric correction of -3.7 , a highly plausible value for a star with T_e in the range of $\sim 2,400$ – $2,700$ K (see refs 6, 11, 34). If we adopt masses of $\sim 0.06 M_{\odot}$ and $\sim 0.08 M_{\odot}$ for VB 8B and VB 8, respectively, and take the orbital separation to be $\geq 9 \times 10^{13}$ cm (ref. 1), we find that the orbital period of the system is ≥ 40 yr. Thus, the properties of this system are entirely consistent with those of a non-interacting binary consisting of a pair of stars whose current masses have not varied from their initial values.

Very low-mass stars have been suggested³⁵ to be a major component of the 'dark matter' that has been inferred, from dynamical considerations^{36,37}, to comprise about half the mass density in the vicinity of the Sun. Estimates of the space density of such stars, based on extrapolation from the observed numbers of low-mass stars that are on the hydrogen-burning main sequence (see, for example, ref. 38), can plausibly account for this 'missing mass'. Detailed studies of the detectability of such objects with ground-based instruments or future space observatories (such as the Hubble Space Telescope and the Space Infrared Telescope Facility) have been carried out by, for example, Staller and de Jong³⁸. The present work has yielded

more reliable cooling curves for very low-mass stars, so that this type of study should be repeated; we are carrying out such investigations and shall report our results elsewhere.

We thank D. McCarthy for a helpful introduction to this subject before publication of his results, and H. Graboske, J. McClintock, R. Probst, D. Stevenson and S. Tremaine for stimulating discussions. L.A.N. acknowledges a postdoctoral fellowship from the NSERC of Canada. This work was supported in part by NASA under contracts NAS5-24441 and NGL-22-009-638 and grant NSG 7643, and by the NSF under grant AST-8217451.

Received 28 February; accepted 13 May 1985.

- McCarthy, D. W. Jr, Probst, R. G. & Low, F. J. *Astrophys. J.* **290**, L9–L13 (1985).
- van Biesbroeck, G. *Astr. J.* **66**, 528–530 (1961).
- Eggen, O. J. *Astrophys. J.* **226**, 405 (1978).
- Walker, A. *Mon. Not. astr. Soc. S. Afr.* **39**, 96–100 (1980).
- Probst, R. G. & Liebert, J. *Astrophys. J.* **274**, 245–251 (1983).
- Hoxie, D. T. *Astrophys. J.* **161**, 1083–1099 (1970).
- Grossman, A. S., Hays, D. & Graboske, H. C. Jr *Astr. Astrophys.* **30**, 95–103 (1974).
- Faber, S. M. & Gallagher, J. S. A. *Rev. Astr. Astrophys.* **17**, 135–187 (1979).
- Carr, B. J., Bond, J. R. & Arnett, W. D. *Astrophys. J.* **277**, 445–469 (1984).
- van de Kamp, P. *Publ. astr. Soc. Pacif.* **81**, 5–10 (1969).
- Popper, D. M. A. *Rev. Astr. Astrophys.* **18**, 115–164 (1980).
- Probst, R. G. *Astr. J.* **82**, 656–661 (1977).
- Kumar, S. S. *Astrophys. J.* **137**, 1121–1125 (1963).
- Rappaport, S. & Joss, P. C. *Astrophys. J.* **283**, 232–240 (1984).
- van de Kamp, P. A. *Rev. Astr. Astrophys.* **13**, 295–333 (1975).
- Lippincott, S. L. *Space Sci. Rev.* **22**, 153–189 (1978).
- Gatewood, G. *Icarus* **27**, 1–12 (1976).
- Paczynski, B. & Sienkiewicz, R. *Astrophys. J. Lett.* **248**, L27–L30 (1981).
- Rappaport, S., Joss, P. C. & Webbink, R. F. *Astrophys. J.* **254**, 616–640 (1982).
- Hayashi, C. & Nakano, T. *Prog. theor. Phys.* **30**, 460–480 (1963).
- Grossman, A. S. *Astrophys. J.* **161**, 619–632 (1970).
- Gautier, D. & Owen, T. *Nature* **302**, 215–218 (1983).
- Grossman, A. S. & Graboske, H. C. *Astrophys. J.* **164**, 475–490 (1971).
- Cox, J. P. & Giuli, R. T. *Principles of Stellar Structure* (Gordon and Breach, New York, 1968).
- Zapolsky, H. S. & Salpeter, E. E. *Astrophys. J.* **158**, 809–813 (1969).
- van der Linden, T. & Staller, R. *Astr. Astrophys.* **118**, 285–288 (1983).
- Stevenson, D. J. *Proc. astr. Soc. Austr.* **3**, 227–229 (1978).
- Alexander, D. R. *Astrophys. J. Suppl.* **29**, 363 (1975).
- Cox, A. N. & Stewart, J. N. *Astrophys. J. Suppl.* **19**, 243–259 (1970).
- Cox, A. N. & Tabor, J. E. *Astrophys. J. Suppl.* **31**, 271–312 (1976).
- Verbunt, F. & Zwaan, C. *Astr. Astrophys.* **100**, L7–L9 (1981).
- Gehrz, R. D., Black, D. C. & Solomon, P. M. *Science* **224**, 823–830 (1984).
- Shapiro, S. L. & Teukolsky, S. A. *Black Holes, White Dwarfs and Neutron Stars* (Wiley, New York, 1983).
- Johnson, H. L. *Astrophys. J.* **141**, 170–176 (1965).
- Oort, J. H. *Galactic Structure Vol. V* (eds Blaauw, A. & Schmidt, M.) 455–512 (University of Chicago Press, 1965).
- Oort, J. H. *Bull. astr. Inst. Neth.* **15**, 45 (1960).
- Bahcall, J. N. *Astrophys. J.* **287**, 926–944 (1984).
- Staller, R. F. A. & de Jong, T. *Astr. Astrophys.* **98**, 140–148 (1981).

Table 1 Models for VB 8 and VB 8B

M ($M_{\odot}$)	R ($R_{\odot}$)	T_e (K)	L ($10^{-4} L_{\odot}$)	t (10^9 yr)	ρ_c (g cm^{-3})	T_c (10^6 K)	D
VB 8B							
0.01	0.130	1,360	0.52	0.07	40	0.49	1.9
0.02	0.103	1,360	0.33	0.12	150	0.71	2.9
0.03	0.091	1,360	0.26	0.31	330	0.87	3.7
0.04	0.083	1,360	0.21	0.65	570	1.02	4.5
0.05	0.078	1,360	0.19	1.16	880	1.13	5.2
0.06	0.074	1,360	0.17	1.87	1,240	1.25	5.9
0.07	0.070	1,360	0.15	2.94	1,660	1.36	6.5
0.08	0.068	1,360	0.14	5.16	2,140	1.46	7.1
0.085	0.067	1,360	0.13	22.90	2,380	1.53	7.2
VB 8							
0.02	0.131	2,400	5.14	0.03	70	0.85	1.8
0.03	0.110	2,400	3.62	0.07	180	1.18	2.2
0.04	0.098	2,400	2.87	0.14	350	1.46	2.5
0.05	0.090	2,400	2.45	0.27	560	1.73	2.8
0.06	0.085	2,400	2.16	0.47	820	1.97	3.1
0.07	0.080	2,400	1.92	0.78	1,140	2.20	3.4
0.08	0.076	2,400	1.73	1.68	1,490	2.42	3.7
0.085	0.074	2,400	1.64	18.70	1,730	2.46	3.9

The column headings are mass (M), radius (R), effective temperature (T_e), bolometric luminosity (L), age (t), central density (ρ_c), central temperature (T_c), and ratio (D) of total pressure to thermal pressure in the stellar interior.

A radio jet associated with the supernova remnant G332.4+0.1 (Kes 32)

R. S. Roger, D. K. Milne, M. J. Kesteven, R. F. Haynes & K. J. Wellington

Division of Radiophysics, CSIRO, PO Box 76, Epping, New South Wales 2121, Australia

We report here new observations at 843 MHz of the galactic supernova remnant (SNR) G332.4+0.1 (Kes 32). They reveal what appears to be a fairly typical SNR, with a 'jet-like feature' emerging from it and expanding into an extensive 'plume' of low brightness emission. It is as though matter confined within the main body of the remnant has broken out through a nozzle and expanded into the plume. However, measurements at 8.4 GHz show the narrow part of the jet to have a relatively flat spectrum and no detectable polarization. Hence, if actually associated with the remnant, the feature is probably either a jet of particles from a central stellar remnant or a pre-existing gas structure ionized by the progenitor star.

The observations of G332.4+0.1 (Kes 32)¹ form part of a high-resolution, high-sensitivity study of SNR morphology using the University of Sydney Molonglo Observatory synthesis telescope (MOST)². At the declination of G332.4 ($\sim 60^\circ$) the telescope has a synthesized half-power beamwidth of 44 arc s (east–west) by 57 arc s (north–south). The effects of a -8%

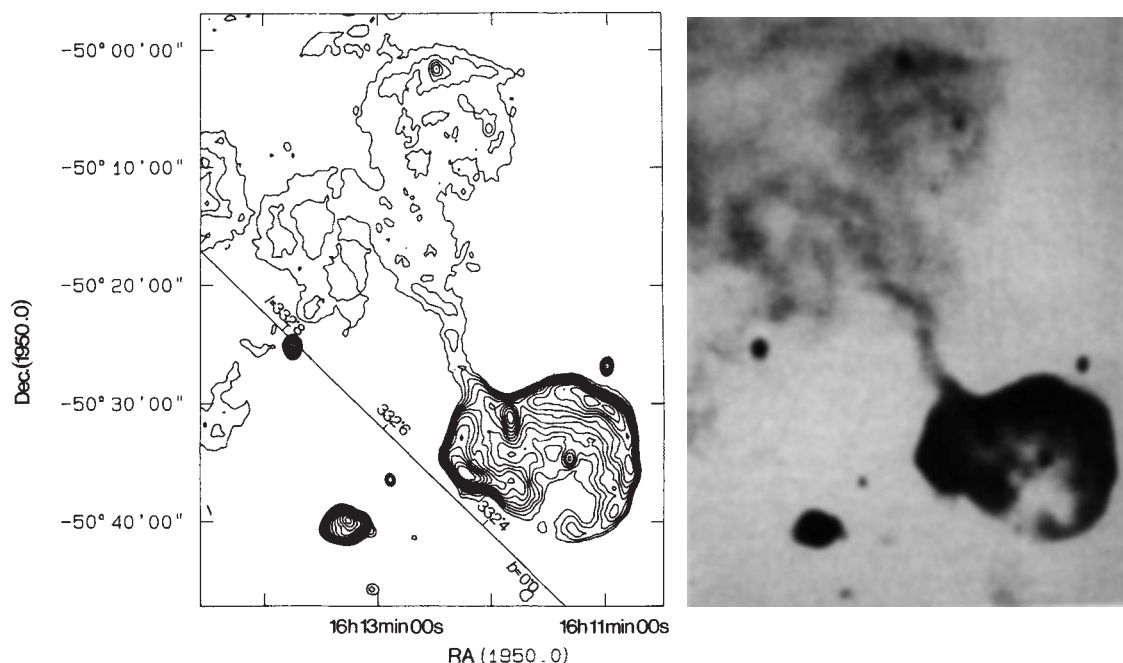


Fig. 1 Contour and grey-scale representations of the 843-MHz radio emission from the SNR G332.4+0.1 (Kes 32) and the jet-like feature emerging from the north-east side of the remnant. Contour levels are: 38, 51, 65, 81, 100, 120, 150, 190, 230, 280, 340, 410 and 500 mJy per beam area.

sidelobe, intrinsic to the mapping process, are removed by cleaning³.

Figure 1 shows contour and grey-scale maps of the 843 MHz emission from the region of G332.4. The remnant is obviously a shell source—albeit irregular in outline and in emission intensity. The two brightest parts of the shell occur on the sides of lowest and highest galactic latitudes. The west rim of the SNR has an almost circular outline, with a weak point source near its centre. The east rim is less regular and the shell appears to have a small radius.

From a point of relatively weak rim emission on the north-east side of the SNR we see a jet-like feature emerging at an angle of $\sim 20^\circ$ to the galactic plane. Near the remnant the 'jet' is narrow, kinked and most intense, whereas further from the remnant the feature broadens to form a 'plume'. At the lowest levels of emission it is difficult to discern where the plume begins to blend with unrelated neighbouring emission.

Complementary observations at 8.4 GHz with the Parkes radio telescope (3 arc min beam) (Fig. 2) show the SNR to have a similar structure to that observed at 843 MHz, with a non-thermal spectral index, which varies over the source in the range -0.45 to -0.55 with an estimated error of 0.05. These measurements, also show the emission from the north-west and south-east rims to be linearly polarized at 8% and 3.5% respectively and at position angles near 135° and 0° . The 8.4-GHz observations extend over only the narrow part of the jet-like feature. A comparison of the two maps shows this part to have a spectral index of -0.05 ± 0.1 ; no polarization could be detected to a limit of $\sim 4\%$.

The distance to G332.4 is uncertain. A lower limit is suggested by the fact that it has no detectable optical or X-ray emission⁴, whereas RCW 103, a SNR some 30 arc min away is detected in these bands⁴; the distance to RCW 103 has been set at 3 kpc (ref. 5). An upper limit of ~ 7 kpc comes from a weak OH absorption⁶, and from the 'surface-brightness-diameter' relation⁷. We have adopted 5 kpc for the calculations that follow. At this distance the diameter of the SNR is ~ 25 pc, and its age, based on the standard Sedov expansion relations, is $\sim 5,000$ yr. On the assumption that the jet and SNR are associated, the narrow part of the jet is ~ 16 pc in length and ~ 3 pc in width. The projected distance from the SNR to the top of the plume is ~ 35 pc.

The association of the jet-like feature with the SNR rests most

strongly on the fact that its emission is strongest and narrowest at the point where it meets the edge of the remnant. The appearance of the jet strongly suggests a nozzle, with material from the interior of the remnant streaming out. However, we have misgivings about such an interpretation, because of the flat spectrum and lack of polarization in the jet, and because of the theoretical difficulty in constructing the rigid skin necessary for the nozzle formation. We have therefore explored two alternative models.

In the first model, we postulate that the jet and plume may comprise synchrotron radiating particles similar in origin and energy spectrum to those which make up the centrally-bright (Crab-like) components of SNRs such as CTB 80 (ref. 8) and G326.3-1.8 (ref. 9). The brightness of the jet implies minimum (equipartition) energy densities of the order of 10^{-10} erg cm⁻³ in the narrow section and a factor of 10 lower in the extended plume. The energetics of the jet do not appear to be a problem, but the collimation may be, given the small opening angle. One possibility could be a supercritical (thick) accretion disk surrounding the central source (see ref. 10); in an alternative scheme, proposed by Benford¹¹, the pulsar provides the collimation. We note, however, that there is no suggestion of a counter jet, whereas both these schemes are inherently symmetric. Further, because of the kinked nature of the jet, further external confinement seems likely. This may be provided by static pressure in the surrounding medium or by self-generated azimuthal magnetic fields, as suggested for the jet in the Crab nebula¹¹. Broadening of the jet and formation of the plume would happen at the point where confinement breaks down owing to the interaction of the flow with the surrounding medium.

The relatively flat spectrum in the narrow part of the jet would suggest that the relativistic particles are recent in origin and, like the centrally bright components of SNRs such as G326.3-1.8, have suffered little differential loss of higher-energy particles. This short time scale implies that velocities of the order of 10^4 km s⁻¹ in the jet itself are required for this material to be transported to the distances observed.

As a second hypothesis, we suggest that the feature might be thermal material. Here we envisage a neutral gas cloud, similar to those seen as extended, narrow (probably columnar) dark clouds, that has been ionized by the pre-supernova. If stellar UV emission were the only source of ionizing radiation, one

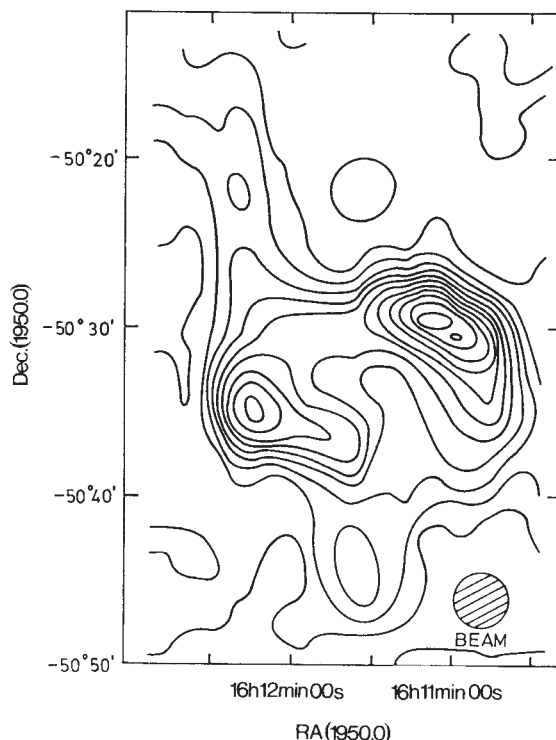


Fig. 2 Contours of the 8.4-GHz radio emission from G332.4+0.1. The contour levels are in increments of 105 mJy per beam area.

might expect recombination to take place after the SN event, with a typical recombination time of $\sim 3,000$ yr. However, the post-SN X-ray flux might be sufficient to significantly delay recombination. The stability of such a structure over the main-sequence lifetime of the ionizing star must be considered. A gas column originally in equilibrium with the surrounding interstellar medium in its cool neutral state could be expected to disperse when ionized on a time scale of $\sim 10^5$ yr.

We have also considered the possibility that the structure might have been shock-ionized following a breakout of the blast wave at a weak point on the shock boundary of the SNR. Such a model¹² has been advanced to explain the filaments which extend far beyond the apparent boundary of IC443. However, the shock would have to advance within the column at a mean velocity of at least four times the velocity of the main remnant shock, and this we view as unlikely.

Features associated with at least five other SNRs have been described as jets. On the smallest scale are the oppositely directed precessing radio jets associated with SS433 within the remnant W50. The radio emission is variable and extends ~ 0.03 pc from the central source in directions aligned with lobes of W50 (ref. 13). It is $\sim 10\%$ linearly polarized and has a spectral index near -0.6 .

A limb-brightened optical protrusion extending about 1 pc from the northern edge of the Crab nebula¹⁴ is most apparent in O III emission¹⁵. A radio counterpart has recently been detected¹⁶ at a level below 1% of the peak brightness of the nebula. This very weak radio feature is highly polarized and is thought to be non-thermal.

On a larger scale, two other remnants, CTB 37A (ref. 17) and CTB 80 (ref. 18), display radio features which we tentatively describe as jets. In the case of CTB 80, arc-like emission can be traced from a central bright core to an estimated distance of 30 pc from the remnant. In CTB 37A, the extended feature is aligned at right angles to the bright shell of the remnant. It is detached from the remnant proper and has the appearance of a misplaced section of shell. In both cases, the spectrum of the emission is distinctly non-thermal with an index not greatly different to that of the SNR.

None of these four jets closely resembles the feature near G332.4.

However, there is a feature associated with the SNR G5.3-1.0 (Milne 56)^{19,20}, which is similar in appearance to the jet and plume described here. The G5.3 jet appears to extend from a compact object out to a distance of about one source diameter, where it fans out into an immense plume. The jet shows no polarization. The plume in G5.3 is discernible to a much greater relative distance than the plume associated with G332.4, which is lost in the nearby confusion. We believe that the two objects fall in the same (SNR) category, and that, contrary to the suggestion of Helfand and Becker²¹, there is no need to invoke a new class of radio source.

Further high-resolution observations of the continuum polarization and of possible recombination line emission for the jet and plume apparently emanating from G332.4+0.1 are required to determine the emission process. New attempts to detect low-level X-ray emission, from either the jet or the postulated central source, would also be valuable.

The Molonglo Observatory is funded by the Australian Research Grants Committee, the University of Sydney Research Committee, and the Science Foundation for Physics within the university. We thank Professors B. Y. Mills and the late A. G. Little, and Dr J. M. Durdin for assistance in using the telescope also Professor D. B. Melrose and Dr G. V. Bicknell for useful discussions.

Received 11 March; accepted 14 May 1985.

1. Kesteven, M. J. L. *Aust. J. Phys.* **21**, 369-376 (1968).
2. Mills, B. Y. *Proc. astr. Soc. Aust.* **4**, 156-159 (1981).
3. Roger, R. S., Milne, D. K., Wellington, K. J., Haynes, R. F. & Kesteven, M. J. *Proc. astr. Soc. Aust.* **5**, 561-562 (1984).
4. Lamb, R. C. & Marker, T. H. *Astrophys. J.* **244**, 94-101 (1981).
5. Caswell, J. L., Murray, J. D., Roger, R. S., Cole, D. J. & Cooke, D. J. *Astr. Astrophys.* **45**, 239-258 (1975).
6. Caswell, J. L. & Haynes, R. F. *Mon. Not. R. astr. Soc.* **173**, 649-669 (1975).
7. Milne, D. K. *Aust. J. Phys.* **32**, 83-89 (1979).
8. Strom, R. C., Angerhofer, P. E. & Dickel, J. R. *Astr. Astrophys.* **139**, 43-49 (1984).
9. Clark, D. H., Caswell, J. L. & Green, A. J. *Aust. J. Phys. Astrophys. Suppl.* No. 37, 75-86 (1975).
10. Calvani, M. & Nobili, L. *Astrophysical Jets* (eds Ferrari, A. & Pacholczyk, A. G.) 189 (Reidel, Dordrecht, 1983).
11. Benford, G. *Astrophys. J.* **282**, 154-160 (1984).
12. Fesen, R. A. *Astrophys. J.* **281**, 658-664 (1984).
13. Seaquist, E. R. *Vistas Astr.* **25**, 79-85 (1981).
14. van den Bergh, S. *Astrophys. J. Lett.* **160**, L27 (1970).
15. Gull, T. R. & Fesen, R. A. *Astrophys. J. Lett.* **260**, L75-78 (1982).
16. Velusamy, T. *Nature* **308**, 251-252 (1984).
17. Milne, D. K. *et al. Mon. Not. R. astr. Soc.* **188**, 437-444 (1979).
18. Angerhofer, P. E., Strom, R. G., Velusamy, T. & Kundu, M. R. *Astr. Astrophys.* **94**, 313-316 (1981).
19. Milne, D. K. & Dickel, J. D. *Nature phys. Sci.* **231**, 33-35 (1971).
20. Becker, R. H. & Helfand, D. J. *Nature* **313**, 115-118 (1985).
21. Helfand, D. J. & Becker, R. H. *Nature* **313**, 118-119 (1985).

Spottiness in the large-scale structure of the microwave background

V. N. Lukash & I. D. Novikov

Space Research Institute, Academy of Sciences of the USSR, Moscow 117810, USSR

Barrow *et al.*¹ have recently considered the large-scale temperature variation ($\Delta T/T$) distribution of the microwave background in the open universe. Interest in this topic has been stimulated by new observations from balloons and the satellite Prognoz. Barrow *et al.*¹ seem to have misunderstood the theory that we have developed²⁻⁴ and arrived at conclusions directly contradicting our results. Here we argue that the main conclusions reached by Barrow *et al.*¹—that concerning the relationship between inhomogeneous and the most general homogeneous anisotropic cosmologies, and their other claim that the sky $\Delta T/T$ pattern of infinite-wavelength inhomogeneities is a superposition of Bianchi type VII models—are erroneous.

The angular pattern of $\Delta T/T$ over the celestial sphere has been called a 'spotty structure' due to the fact that the general form of the perturbation can be represented as a kind of Fourier integral. The basic functions of this expansion are akin to plane waves, and each of them produces a 'spot' (an annulus) in the $\Delta T/T$ pattern of angular size $\theta \sim \Omega = \rho/\rho_{cr}$ which is determined by the space curvature of the universe at the time of observation² (ρ is the total matter density). (In a spatially flat universe ($\Omega = 1$), the 'spot' is a quadrupole structure.) An important feature of ref. 2, distinguishing it from those discussed in the 1950s and 1960s³⁻⁷, is its close relationship to the geometry of homogeneous Bianchi type VII and V models. In this sense, the theory from ref. 2 is a generalization of the spot effect⁸⁻¹⁰ inherent in the Bianchi models mentioned above.

Arbitrary inhomogeneous transverse perturbations, which do not perturb the matter density, are directly related to the homogeneous anisotropic models. For these perturbations, one can always choose plane waves that are circularly polarized and consequently, permit the type VII group of motions. In this way, "... arbitrary vortex and gravitational-wave perturbations can be Fourier expanded into the homogeneous modes of type VII ...".¹² So, the celestial $\Delta T/T$ pattern of transverse perturbations of any scale is a linear superposition of Bianchi type VII spots. Just the opposite occurs with scalar (density) perturbations, the situation Barrow *et al.*¹ deal with.

Scalar waves are nonhomogeneous due to their longitudinal nature. That is why it is, in principle, not possible to relate density perturbations (even of large scale) to homogeneous models of any kind. This directly contradicts the claim by Barrow *et al.* that "... it is possible to model inhomogeneous universes close to the Friedman model as superpositions of the most general homogeneous Bianchi models ...". Nevertheless, the sky pattern of arbitrary scalar inhomogeneities is again a superposition of spots (but nonhomogeneous ones)²⁻⁴. The structure of these nonhomogeneous spots differs from the structure of spots in type VII models.

The angular size of a spot produced by a single scalar plane wave depends on the background space curvature only, and is independent of the perturbation scale, cosmological evolution, recombination dynamics, reheating, missing mass parameters and so on. (These factors determine the fine angular structure and the amplitude of the spot.) The large-scale shape of any spot is unique for all plane waves and can be easily derived using Lobachevski three-geometry (the space geometry of the open universe)^{4,10,11}:

$$\left| \frac{\Delta T}{T} \right| \sim \left(\frac{2\pi \tan \theta/2}{1 + (\kappa \tan \theta/2)^2} \right)^2 \quad (1)$$

Here θ is the angle between the wave-vector and the line-of-sight, $\kappa = \exp(-h\chi_r)$, h^{-1} is the curvature radius of the Lobachevski space, $\chi_r = \int_{t_0}^t dt/a$ is the geodesic three-distance of the observer (t_0 is from the recombination sphere (t_r), $a = a(t)$ is the scale factor. For the euclidian space ($h = 0$) and equation (1) reduces to quadrupole dependence ($\sim \sin^2 \theta$). For the Lobachevski space $\kappa < 1$ and all $\Delta T/T$ anisotropy develops inside the annulus of angular radius $\theta_0 = 2 \arctan \kappa$ and width $\sim \theta_0$. For small Ω , $\kappa \approx \Omega/4$ and the angular size of the spot is Ω . The large-scale profile, equation (1), of a single spot may already serve as an illustration of the effect of spottiness in open universes: for $\Omega = 1$ ($h = 0$) an arbitrary superposition of spots always results in a quadrupole distribution, while for $\Omega < 1$ ($h \neq 0$) a statistical distribution of $\Delta T/T$ with an angular correlation scale $\sim \Omega$ takes place. We now consider the fine structure of non-homogeneous spots.

The value of the wave-vector modulus k determines the fine structure of a spot, which actually consists of a number of narrow concentric rings packed inside the spot annulus. For $k > h$, the large-scale spot profile, equation (1), modulates the sequence of these rings. With decreasing k the rings inside the annulus smear out (the width of every ring increases) and the overall spot profile approaches equation (1) as $k \rightarrow 0$. The spot amplitude can be of any sign, so cold- and hotspots have equal probabilities.

In the infinite-wavelength limit ($k = 0$), plane waves do not perturb space curvature and become homogeneous—they exhibit the translational symmetry of type V models. [This is valid for all plane waves (not only for scalar ones). Recall that for the limit $k \rightarrow 0$, Bianchi type VII group of motions becomes type V. From the physical point of view, each of these models consists of three plane waves on the Friedman background with the wave vectors in the same direction: the scalar one with the infinite wavelength plus the vortex and the gravitational ones. The only difference between the two models is that in the type VII model the latter two waves have finite wavelengths (and are circularly polarized), while in the type V model they are infinite¹¹.] Thus, as was formulated earlier in refs 2-4, any infinite-scale perturbation is a superposition of type V models. This contradicts another conclusion of Barrow *et al.*¹ that infinite-wavelength scalar inhomogeneities are the superpositions of type VII models. The difference in conclusions means that the true cause of disagreement stems from the initial basic assumptions.

So, to analyse nonhomogeneous perturbations, we incorporate the Fourier expansion into the plane waves (which reduces to the well-known Fourier integral over e^{ikx} in the euclidian universe). Barrow *et al.*, in contrast, use the method of covering (piecewise smooth approximation) a nonhomogeneous solution with pieces (patches) of homogeneous models. They have developed their conception on the basis of an evolution formula for the density perturbations $\delta\rho/\rho$ in a dust-like medium (see their equations (16) and (17)). Such perturbations evolve in the same way for all wavelengths. Analogous expressions (with arbitrary functions of spatial coordinates replaced by constants) also exist in homogeneous models (see their equations (6) and (15)) which according to Barrow *et al.* enables one to approximate large-scale density perturbations in its most general form by "... an ensemble of homogeneous Bianchi VII models of differing density ...", that is with varying constants. Such an approximation over the entire spectrum is completely unjustified because homogeneous density perturbations originate from scalar waves with infinite wavelengths. At the same time, to invoke vortex and gravitational-wave modes (which perturb the curvature in type VII models) to calculate the shear evolution (see their equations (6) and (15)) is physically meaningless because we consider the density perturbations only. On the other hand, from the full Fourier integral, one infers that, when covering a large-scale density inhomogeneity with homogeneous patches of finite sizes, the dominant contribution will be due to the essentially nonhomogeneous modes with wavelengths of the order of the inhomogeneity size. The approximation of a non-homogeneous solution with homogeneous ones is justified in the limit of infinite-wavelength perturbations only, but then it requires only type V models.

The analogy in the behaviour of homogeneous and non-homogeneous scalar perturbations, based on the invariant evolution law for $\delta\rho/\rho$, cannot be extended to non-dust-like media, while the effect of spottiness is always there. It results from the global curvature of the universe ($\Omega < 1$)^{2,8} and depends only quantitatively on details of evolution.

We conclude by emphasizing that in the most general case, the spotty structure of $\Delta T/T$ in open universes is described by a Fourier integral and cannot, in principle, be reduced to an expansion in homogeneous Bianchi models. The discovery of a 'spotty' structure (or of its absence) of the microwave background would be extremely important for cosmology. It could answer basic cosmological questions such as what was the spectrum of primordial perturbations and what is the total density of all kinds of cosmic matter (including its invisible components).

Received 18 June 1984; accepted 19 April 1985.

1. Barrow, J. D., Juszkiewicz, R. & Sonoda, D. H. *Nature* **305**, 397-402 (1983).
2. Lukash, V. N. *Abstr. Contr. Pap., 8th GRG int. Conf.*, Waterloo, 237 (Univ. Waterloo, 1977).
3. Bisnovaty-Kogan, G. S., Lukash, V. N. & Novikov, I. D. in *Proc. 5th Regional Meet. Astronomy*, Liege, G.1.1-G.1.20 (IAU/EPS, 1980).
4. Lukash, V. N. *Proc. IAU Symp. No. 104*, 149-152 (1982); thesis, Space Res. Inst., Moscow (1983).

5. Shapiro, I. S. *Dokl. Acad. Nauk. SSSR* **106**, 647-652 (1956); *Soviet Phys. Dokl.* **1**, 91-94 (1956); *Phys. Lett.* **1**, 253-255 (1962).
6. Gelfand, I. M. & Graev, M. I. *Proc. Moscow Math. Soc.* **11**, 243-286 (1962).
7. Vilenkin, N. Ya. & Smorodinskii, Ya. A. *JETP* **46**, 1793-1808 (1964); *Soviet Phys. JETP* **19**, 1209-1218 (1964).
8. Novikov, I. D. *Astr. Zh.* **45**, 538-540 (1968); *Soviet Astr.* **12**, 427-429 (1968).
9. Grichshuk, L. P., Doroshkevich, A. G. & Novikov, I. D. *JEPT* **55**, 2281-2290 (1968); *Soviet Phys. JETP* **28**, 1210-1216 (1969).
10. Doroshkevich, A. G., Lukash, V. N. & Novikov, I. D. *Astr. Zh.* **51**, 940-950 (1974); *Soviet Astr.* **18**, 554-560 (1974); *JETP* **64**, 1457-1474 (1973); *Soviet Phys. JETP* **37**, 739-746 (1973).
11. Lukash, V. N. *JETP* **67**, 1594-1608 (1974); *Soviet Phys. JETP* **40**, 792-799 (1975); *Nuovo Cim.* **35B**, 268-292 (1976); thesis, Phys. Technic Inst., Moscow (1974).

John D. Barrow*, R. Juszkiewicz† & D. H. Sonoda* reply

* Astronomy Centre, University of Sussex, Brighton BN1 9QH, UK

† Copernicus Astronomical Center, Bartycka 18, 00716 Warsaw, Poland

Lukash and Novikov have given an interesting survey of their analyses of some of the features that may exist in the microwave background radiation in open universes. Here, we answer some criticisms that they make of our earlier analysis¹, and point out a further unusual feature that appears in the temperature distribution of the most general flat and open cosmological models.

Lukash and Novikov claim that in ref. 1 we erroneously represented infinite wavelength density perturbations by ensembles of Bianchi type VII universes at linear order, rather than by cosmological models of Bianchi types I or V, as required. This claim is incorrect. We found, and stated, that infinite-wavelength, pressure-free density inhomogeneities could be represented by ensembles of VII models only when one of their defining parameters, $x = h^{1/2}(1 - \Omega_0)^{-1/2}$, tends to infinity. In this limit, the type VII_h and VII₀ models specialize identically to the type V and I models respectively; here, Ω_0 is the density of the Universe today in units of the critical density and x or h is an arbitrary constant once Ω_0 is specified. (In the VII₀ case, the parameter x still exists when the limits $h \rightarrow 0$ and $\Omega_0 \rightarrow 1$ are taken simultaneously; see ref. 2 for details). Accordingly, all computations in ref. 1 of the effects of inhomogeneity on the background radiation were performed using the type I and V models, as Lukash and Novikov argue is necessary.

Finally, we would like to comment on one important difference between the hotspot patterns that arise in open universes of type VII and those arising in open universes of type V which were described by Lukash and Novikov and by ourselves¹. Although the type VII models are not relevant for the description of density inhomogeneities, they can describe the effects of gravitational waves and vorticity on the temperature anisotropy of the microwave background.

The general equation used by Lukash and Novikov to describe the hotspot structure is based on the type V paradigm where the spatial curvature is isotropic, and does not completely describe the most general hotspots generated by vortical and gravitational wave distortions; these will be described by the type VII_h behaviour rather than the simpler type V case. The geodesics in open type V universes exhibit focusing of a quadrupole temperature profile into a hotspot feature but with no change in the azimuthal coordinate of the geodesics propagating from their last scattering to the observer today. However, in type VII models there arises a geodesic spiralling effect which is superimposed on the hotspot structure. This feature was first noted by Collins and Hawking³ and its observational effects are examined in detail in ref. 2. Figure 1 shows examples of the temperature anisotropy patterns predicted in VII₀ and VII_h universes. In the VII₀ case, where $\Omega_0 = 1$, there is a pure quadrupole plus the geodesic spiralling effect, whereas in an open VII_h model with $\Omega_0 = 0.7$, we see a spiral plus quadrupole pattern focused into a spiral hotspot. The number of turns of the spiral is inversely proportional to the parameter x . In the limit $x \rightarrow \infty$, in which case the spatial curvature becomes isotropic, the VII₀ and VII_h models become those of types I and V again, the spiral effect disappears and we recover either a quadrupole or the

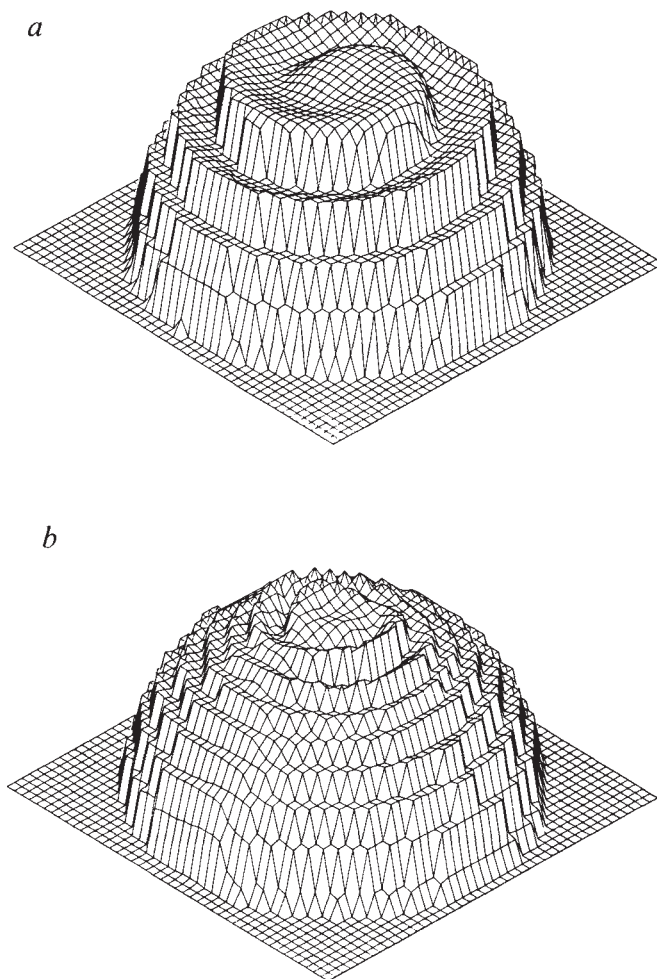


Fig. 1 Microwave background temperature profiles, $T(\theta_0, \phi_0)$, predicted in the region of observed angles $0 < \phi_0 < 2\pi$, $\pi/2 \leq \theta_0 < \pi$. The radial distance represents the magnitude of $T(\theta_0, \phi_0)$. We assume that the last scattering of the photons occurred at a redshift of 1,000, and the anisotropy amplitude is taken to be the maximum allowed by current observations. *a*, Spatially homogeneous-type VII₀ cosmological model with $x = 0.067$ and $\Omega_0 = 1$. *b*, An open spatially homogeneous cosmological model of type VII_h with $x = 0.067$ and $\Omega_0 = 0.7$. In *a*, the pattern is approximately a pure quadrupole modulated by the geodesic spiralling effect. In *b*, both the quadrupole and spiral patterns seen in *a* have been focused into a region of smaller angular scale towards the axis $\theta_0 = \pi$ by the negative spatial curvature. The result is a spiral hotspot. The spirals are left-handed in the observed angles. As $x \rightarrow \infty$ the spiral effect disappears and the hotspot in *b* degenerates into the simple variety found in the type V models, and *a* becomes the pure quadrupole of type I. For further details and discussion see ref. 2.

simple hotspot structure discussed in the preceding letter and ref. 1. Note that because this spiral feature does not arise in closed universes, it may enable observers to determine whether the Universe is open or closed, no matter how close its density lies to the critical value².

We thank J. Stein-Schabes and N. Kaiser for comments and V. Lukash and I. Novikov for communicating their results before publication. D.H.S. and R.J. acknowledge financial support at Sussex from the SERC.

Received 18 June 1984; accepted 19 April 1985.

1. Barrow, J. D., Juszkiewicz, R. & Sonoda, D. H. *Nature* **305**, 397-402 (1983).
2. Barrow, J. D., Juszkiewicz, R. & Sonoda, D. H. *Mon. Not. R. astr. Soc.* **213**, 917-943 (1985).
3. Collins, C. B. & Hawking, S. W. *Mon. Not. R. astr. Soc.* **162**, 307-320 (1973).

A search for evidence of nuclearites in astrophysical pulse experiments

N. A. Porter*, D. J. Fegan*, G. C. MacNeill* & T. C. Weekes†

* Physics Department, University College, Dublin 4, Ireland

† Harvard-Smithsonian Center for Astrophysics, Whipple Observatory, Arizona 85645-0097, USA

De Rújula and Glashow¹ have suggested that nuclearites, aggregates of up, down and strange quarks in roughly equal proportions, may form a component of the material reaching the Earth from the Galaxy. On traversing the atmosphere they will look something like meteors, but will travel faster. The masses of these aggregates may vary over a wide range and their velocities will be typically 250 km s^{-1} , corresponding to the Sun's galactic rotation. If all the dark matter in the Galaxy is assumed to be in the form of nuclearites, a limit can be set to the incoming flux. We report here upper limits derived from four experiments, originally carried out to detect cosmic γ rays, as well as some derived by other authors, which would be sensitive to pulses of light from nuclearites in the lower atmosphere. Our upper limits are compatible with the suggested maximum flux and competitive with a search for tracks etched in mica², which could also be caused by nuclearites.

De Rújula and Glashow show that the visual magnitude of an atmospheric nuclearite:

$$M = 10.8 - 1.67 \log_{10} (m/10^{-6}) + 5 \log_{10} (h/10 \text{ km}) \quad (1)$$

where m is the mass of the nuclearite in grams and h its distance from the detector. The nuclearite will radiate predominantly in the lower atmosphere, effectively below:

$$h_{\text{max}} = 2.7 \text{ km} \ln (m/1.2 \times 10^{-5} \text{ g}) \quad (2)$$

In our experiments, m ranged from 10^{-3} to 10 g , and h_{max} was typically 20–30 km, except in one case where it was constrained by the arrangement of the optical system.

The first five experiments that we consider were carried out at the Whipple Observatory, Arizona. The first³, in 1969–72, used the 10-m reflector to search for γ rays from the Crab Nebula or, for pulsars in the 10^{11} -eV region. A total of 112 h of observation were taken at various zenith angles. To trigger the system, ~ 50 photoelectrons were required in an integration time of 10^{-8} s . In a subsequent analysis⁴, a search was made in the data for bursts of events lasting up to 0.1 s . This analysis is the basis of the present discussion; 12 events were required to constitute a burst. The transit time through the field at a typical detectable height was $\sim 10^{-2} \text{ s}$. To produce 12 events from random fluctuations in 10^{-2} s , with a basic integration time of 10^{-8} s , requires about a 4.5σ upward deviation from the mean light of the nuclearite in these 10^{-8} s intervals. (The night-sky background contribution⁵ was small, $\sim 10\%$ of that from the nuclearite). Fifty photoelectrons are needed, σ therefore is 11.1, and $\sigma^2 = 123$, the mean number of photoelectrons per 10^{-8} s interval. With a 15% photocathode efficiency at the photomulti-

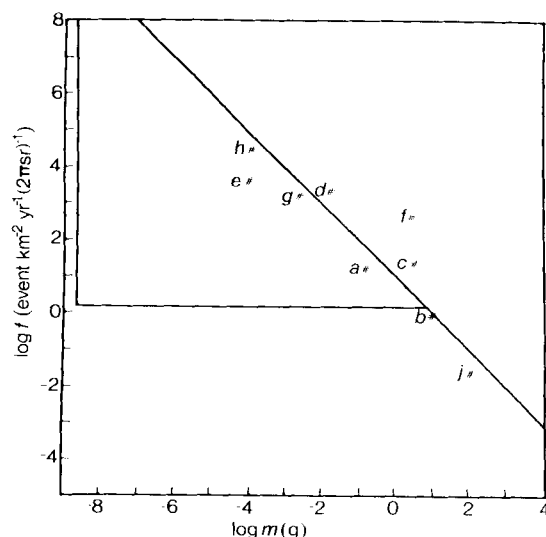


Fig. 1 Mass of the nuclearites plotted against incoming flux. The sources are detailed in Table 1.

plier, we obtain a photon number of 823 photons in 10^{-8} s average illumination over the 10^{-2} s transit time. The area of the reflector was effectively 60 m^2 , leading to a flux of 1.37×10^{-3} photons cm^{-2} in 10^{-8} s , or 1.37×10^5 photons $\text{cm}^{-2} \text{ s}^{-1}$. Converting this to $\text{erg cm}^{-2} \text{ s}^{-1}$ to obtain the visual magnitude⁶ and taking $0 \text{ mag} = 8.5 \times 10^{-6} \text{ erg cm}^{-2} \text{ s}^{-1}$, we obtain a limiting magnitude of 3. Combining equations (1) and (2) the mass became 0.16 g , h_{max} was 25 km , and a typical distance h was 15 km .

The surface area of the conical frustum from 2.2 km (altitude of observatory) up to 25 km at a mean zenith angle of 25° and full field of view 1° was 16.6 km^2 . Upward moving nuclearites occur for this mass, so the whole area is effective. The statistical upper limit, at the 99% level, from the burst analysis, was 477 events yr^{-1} . We therefore obtain an upper limit flux of 28.7 events $\text{km}^{-2} \text{ yr}^{-1} (2\pi \text{ sr})^{-1}$, at a mass $m = 0.16 \text{ g}$. This limit is plotted as a in Fig. 1, the diagonal line is the upper limit to the cosmic flux, derived by De Rújula and Glashow on the assumption that the entire dark matter of the Galaxy consists of nuclearites. The triangular area in Fig. 1 represents the limits derived in ref. 1 from an experiment by Price *et al.*² using the etching of tracks in buried mica samples, originally intended for the detection of monopoles. It will be seen that our limit falls below the maximal flux and is compatible with the mica experiment.

We are aware of eight other experiments which can be similarly analysed. Their main features are summarized in Table 1 and the resultant flux limits plotted in Fig. 1.

The results are generally comparable with those obtained in the mica experiment. Only that of Ogelman¹⁴ can confidently be placed below both the maximal flux and the mica limit. Experiments designed specifically for nuclearite detection could, however, significantly reduce the limits by optical techniques.

Table 1 Cosmic flux limits for nuclearites of various masses

Ref.	Mirror diameter (m)	Mass (g)	Flux limit (events $\text{km}^{-2} \text{ yr}^{-1} (2\pi \text{ sr})^{-1}$)	Point in Fig. 1
7	1.5×2	10	1.15	b
8	1.5×3	3	4.3	c
9	10×1	1.5×10^{-2}	1,700	d
10	10×1	10^{-4}	5,000	e
11	1.5×1	0.2	380	f
12	0.8×1	3×10^{-3}	1,710	g
13	1×1	10^{-4}	3.5×10^4	h
14	None, PM tubes only	250	0.013	j

Received 11 March; accepted 2 May 1985.

- De Rújula, A. & Glashow, S. L. *Nature* **312**, 734–737 (1984).
- Price, P. B. *et al. Phys. Rev. Lett.* **52**, 1265–1268 (1984).
- Helmken, H. F. *et al. Astrophys. J.* **184**, 245–250 (1973).
- Porter, N. A. & Weekes, T. C. *Nature* **277**, 199 (1979).
- Jelley, J. V. *Cerenkov Radiation*, 219 (Pergamon, London, 1958).
- Allen, C. W. *Astrophysical Quantities* 190 (University of London Press, 1964).
- Porter, N. A., Delaney, T., Helmken, H. F. & Weekes, T. C. *Nuovo Cimento* **32B**, 514–518 (1976).
- Porter, N. A. & Weekes, T. C. *Astrophys. J.* **212**, 224–226 (1977).
- Porter, N. A. *et al. Nature* **304**, 606–607 (1983).
- Cook, A. F., Weekes, T. C., Williams, J. T. & O'Mongain, E. *Mon. Not. R. astr. Soc.* **193**, 645–666 (1980).
- Charman, W. N. & Jelley, J. V. *J. Phys. A*, **5**, 773–780 (1972).
- Byrne, P. B. & Wayman, P. A. *Mon. Not. R. astr. Soc.* **173**, 537–552 (1975).
- Baird, G. A., Delaney, T. J. & Lawless, B. G. *Observatory* **92**, 233–234 (1972).
- Ogelman, H. *J. geophys. Res.* **78**, 3033–3039 (1973).

Penrose tiling observed in a quasi-crystal

L. A. Bursill* & Peng Ju Lin

Center for Solid State Science, Arizona State University, Tempe, Arizona 85287, USA

Three-dimensional icosahedral symmetry has been reported¹ for an alloy consisting of aluminium with 14 atomic% manganese. The electron diffraction patterns appear consistent with a three-dimensional Penrose tiling² or quasi-lattice^{2,3}, although one report has suggested a 20-fold multiple twinning of a close-packed structure⁴. Clearly, however, the determination of the reciprocal lattice is only a first step; it is a further problem to determine the atomic motif(s) required to complete the structure. High-resolution electron microscope images of this alloy, reported here, reveal the localized pentagonal pseudo-symmetry of the structure and show that it projects as a two-dimensional Penrose tiling^{2,5}. A structural model is derived, based directly on the experimental images and an analysis of the polyhedra occurring in the crystallographic phase Al_6Mn .

The specimen was provided through the courtesy of Dr Leo Bendersky (National Bureau of Standards, Washington, DC) and was prepared as described in ref. 1. Figure 1 shows a view, looking exactly along one of the 5-fold symmetry axes, of an area considerably less than 100 Å thick. The imaging conditions in the JEOL-200CX instrument used were chosen so that, according to the usual theories of image contrast⁶, we would expect the areas of black contrast to represent regions of relatively high atomic density, at a resolution of ~ 2.4 Å. There is a striking resemblance to the original Penrose drawings of tiling patterns⁷. The enlargement in Fig. 2a shows a pronounced localized 5-fold pseudo-symmetry, the centre of which maps almost perfectly onto one of Penrose's drawings (Fig. 2b). In general, the pentagonal symmetry is short-range, rarely extending further than 15 Å, as expected for Penrose tiling. Rotation of the electron micrograph with respect to a fixed glancing angle reveals five interpenetrating sets of spacings in the ratio $1:\tau$ ($=1+\sqrt{5}/2$), characteristic of the Fibonacci series. Note also the decagonal rings (diameter 14.5 Å) of white spots, pentagonal stars of white spots (separation 5.4 Å) and the decagonal and pentagonal rings of black contrast (diameters 8.8 and 5.4 Å, respectively). These are shown schematically in Fig. 2c, which also allows the five Fibonacci sequences comprising this area to be identified.

Following the idea of Mackay², we tried to build a structural model using acute and obtuse rhombohedral units, taking the rhombic edge to be 2.7 Å (the mean Al-Al separation in crystalline aluminium). Although some features of the image could be matched, for example the decagonal rings of black contrast correspond geometrically to the projected waist of Mackay's rhombic triacontahedra, we could not match the diameters or disposition of the small dark rings on either the larger decagonal rings of white spots or the pentagonal stars of white spots. We conclude that the structure is not based on such slightly distorted elements of cubic close-packing, at least if the white spots represent relatively low densities of atoms. It appears that, in this case, we must search for a different structural motif for the quasi-lattice.

A careful consideration of the structure reported⁷ for Al_6Mn reveals (Fig. 3a) that it may be considered as a stacking of two symmetrically equivalent layers composed of polyhedra which differ from an icosahedron only in that one vertex is unoccupied (Fig. 3a inset). These truncated icosahedra (or TICOS) share edges within layers (Fig. 3b) and corners between layers (Fig. 3c). Note that each TICOS contains: one Mn atom not shared with adjacent polyhedra; eight Al atoms, shared between those TICOS within layers; two Al atoms, shared with one other

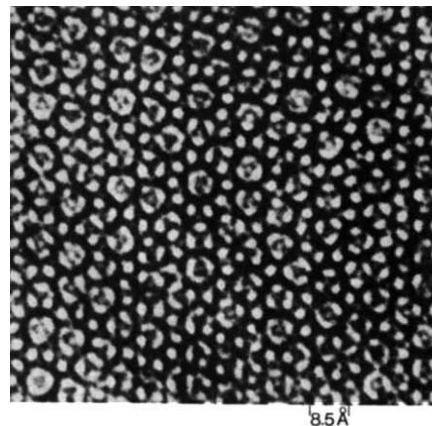


Fig. 1 High-resolution electron microscope image of Al:14 at.% Mn viewed along a 5-fold symmetry axis. Note the localized centres of pentagonal symmetry and absence of translational symmetry. Large-scale and fine-scale Penrose tiling patterns may be traced over.

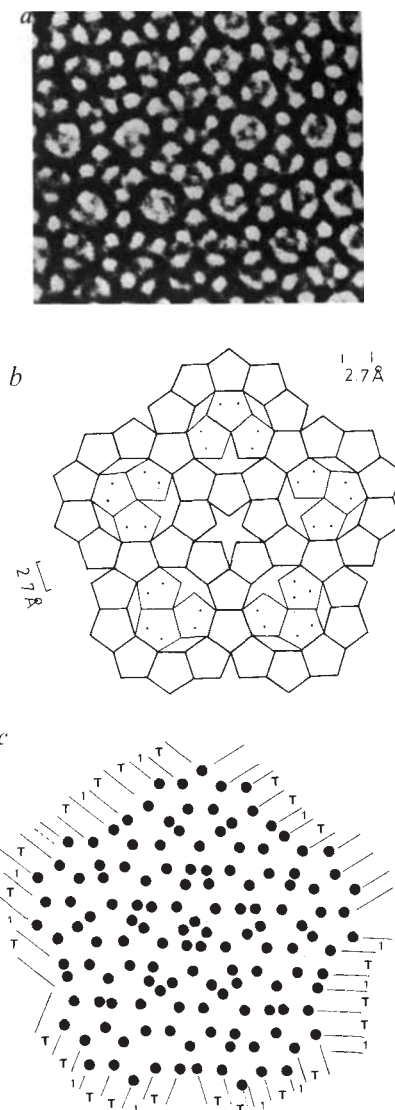


Fig. 2 Enlargement of a pronounced centre of pentagonal pseudo-symmetry (a), together with a copy of an original Penrose drawing, showing almost perfect mapping of white spots (pentagon centres) and dark contrast (lines) (b). c, Schematic drawing showing decagonal and pentagonal rings representing nodes of the Penrose tiling pattern, allowing the interpenetrating Fibonacci series of spacings to be identified.

* Permanent address: School of Physics, University of Melbourne, Parkville, Victoria 3052, Australia.

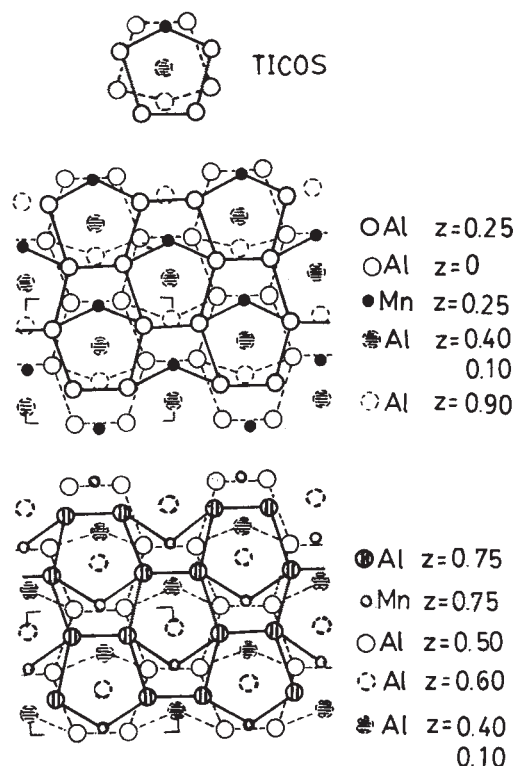


Fig. 3 [001] Projection of crystalline Al_6Mn , showing edge-sharing of truncated icosahedra (TICOS) within layers and corner-sharing between layers. Each TICOS contains one Mn atom (on surface) and one Al atom (internal) which are not shared with adjacent TICOS, the eight remaining Al atoms being shared between two TICOS.

TICOS in adjoining layers; and a further Al atom inside each TICOS, which is therefore unshared. The net result is, of course, a stoichiometry of Al_6Mn .

This analysis suggested that the observed Penrose tiling could be explained by a cooperative rearrangement of TICOS. Figure 4 shows two levels of a model for the central area of Fig. 2. The pentagonal star at the first level is surrounded by decagonal rings composed of rings of five TICOS at both first and second levels, tessellated according to the observed tiling pattern. The model could readily be extended, at least in two dimensions, to match details of the dark (and white) contrast variations in the experimental images. Note that the algorithm used was that a white spot represents the centre of a column of TICOS along the projection axis. Also, there is a relatively low density of Al atoms inside the TICOS, Mn always occupying asymmetrically coordinated sites on the pentagonal rings at the surface of individual TICOS. The stoichiometry Al_6Mn is preserved in this model, since on average each TICOS retains the same number of shared Al atoms (8) and unshared Mn (1) and Al (1) atoms.

The essential icosahedral geometry of the TICOS gives us some confidence, justified so far by model-building, that a fully three-dimensional model may be achieved in which six such interpenetrating two-dimensional networks coexist. Dark-field observations suggest correlation lengths of ~ 50 – 100 Å along the 5-fold axes. Consideration of the configurations of black and white contrasts expected for overlapping in projections of the possible two-dimensional patterns has so far been consistent with the experimental images. We are now investigating whether existing electron dynamical techniques can be used to calculate the images, and it should be possible to calculate images for appropriate sections of the model using existing techniques for non-periodic objects⁸.

The experimental observations have already been extended to both 3-fold and 2-fold axes. Pseudo-symmetric centres are again observed. In addition, stacking faults and other defects occur in the quasi-lattice (sight along the dark fringes in Fig. 1);

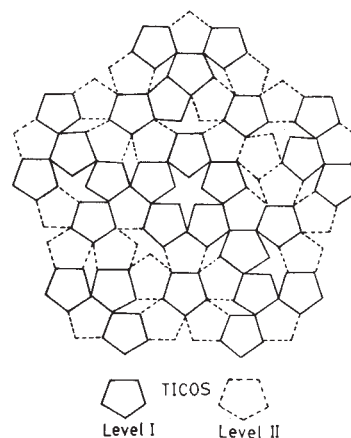


Fig. 4 Assemblage of TICOS polyhedra to generate two-dimensional Penrose tiling reproducing the central area of Fig. 2. TICOS were placed at the centre of each white spot in the experimental image.

localized areas appear to transform to less-ordered (glassy) states and the crystalline phase may also occur (L.A.B. and P.J.L., in preparation).

The structural principles described above, which we believe to be justified reasonably well by both the experimental images, and our analysis of the structure reported for Al_6Mn , should be regarded as tentative. Nevertheless, we hope they will provide a realistic basis for future studies of the alloy's structural and physical properties. What is most intriguing, of course, is whether we are concerned with a material having singular structural properties because of the chemistry of Al and Mn, or whether the principles suggested by the quasi-crystal concept will find more widespread application. We favour the latter possibility.

We thank Professor J. M. Cowley and the Facility for High Resolution Electron Microscopy, in the Center for Solid State Science, Arizona State University, established from the NSF (grant DMR-8306501).

Received 25 March; accepted 24 May 1985.

1. Shechtman, D., Blech, I., Gratias, D. & Cahn, J. W. *Phys. Rev. Lett.* **53**, 1951–1954 (1984).
2. Mackay, A.L. *Physica* **114A**, 609–613 (1982).
3. Levine, D. & Steinhardt, P. *Phys. Rev. Lett.* **53**, 2477–2480 (1984).
4. Field, R. D. & Fraser, H. L. *Mater. Sci. Engng* **68**, L17–L21 (1985).
5. Penrose, R. *Bull. Inst. Math. Appl.* **10**, No. 718, 266–271 (1974); *Eureka* **39**, 16–22 (1978).
6. Cowley, J. M. & Iijima, S. *Z. Naturforsch.* **27A**, 445–451 (1972).
7. Nicol, A. D. I. *Acta crystallogr.* **6**, 285–293 (1953).
8. Barry, J. C. & Bursill, L. A. *Micron* **13**, 249–250 (1982).

Quantization effects in the photoelectrochemistry of superlattice photoelectrodes

A. J. Nozik, B. R. Thacker & J. M. Olson

Solar Energy Research Institute, Golden, Colorado 80401, USA

Size quantization effects have created much recent interest. In solid state physics, for example, the novel effects exhibited by thin layered semiconductor structures in the form of multiple-quantum-wells (MQW) or superlattices (SL) are under intense study^{1–7}. In photoelectrochemistry, quantum-size effects produced by very small (< 100 Å) semiconductor particles have been analysed^{8–11}. We report here the first quantum size effects produced in a photoelectrochemical system using a superlattice photoelectrode. The electrode contained 40 alternating layers of undoped GaAs and $\text{GaAs}_{0.5}\text{P}_{0.5}$ each 250 Å thick. This structure produces a strained-layer superlattice (SLS) with 20 GaAs quantum wells having a well width of 250 Å; the $\text{GaAs}_{0.5}\text{P}_{0.5}$ barrier widths are also 250 Å thick. The barrier heights for electrons in the conduction band

and for holes in the valence band for this SLS are 0.28 eV and 0.30 eV, respectively⁶. The photocurrent–wavelength response shows remarkable structure at room temperature that arises from the presence of discrete energy levels in the GaAs quantum wells. This structure is readily resolvable into multiple photocurrent peaks that are in excellent agreement with theoretical predictions of the discrete energy levels in the quantum wells. These results have important implications for photoelectrochemical energy conversion and for the characterization and analysis of SL and MQW structures in general.

The SLS structure was grown by metal-organic chemical vapour deposition (MOCVD) using continuous flows of trimethylgallium, PH_3 , and AsH_3 . The growth rate and growth temperature were $0.1 \mu\text{m min}^{-1}$ and 750°C , respectively. A $10\text{-}\mu\text{m}$ $\text{GaAs}_{1-x}\text{P}_x$ buffer layer graded from $x = 0$ to 0.25 in five equal steps was inserted between the $\text{p}^+\text{-GaAs}$ substrate and the SLS structure. The $\text{p}^+\text{-GaAs}$ wafers were oriented 2° off (100) towards (110) and were cleaned before growth using a $6:1:1$ $\text{H}_2\text{SO}_4:\text{H}_2\text{O}_2:\text{H}_2\text{O}$ etch. The SLS layers are not lattice-matched and large strains are, therefore, generated in the layers both parallel (biaxial shear strain) and perpendicular (hydrostatic strain) to the layers^{5,6}. A scanning electron micrograph (SEM) of the layered structure is shown in Fig. 1; the large strain results in wavy, but equidistant, layers. The top layer is GaAs. The SLS structure studied in this work has the same configuration as that reported by Gourley and Biefeld⁶ to exhibit quantization effects in optical and photoluminescence data.

The theoretical energy-level scheme for a $250 \text{ \AA}$ $\text{GaAs}/\text{GaAs}_{0.5}\text{P}_{0.5}$ SLS is shown in Fig. 2. The energy-level calculation is a simple modification of that previously reported⁶ and includes the effects of lattice strain. The selection rules for optical transitions in quantum wells require that $\Delta n = 0$, where n is the quantum number of the respective valence and conduction band states. The allowed transitions are indicated in Fig. 2.

Photoelectrodes were prepared from the SLS wafer by contacting the $\text{p}^+\text{-GaAs}$ substrate with a gold ohmic contact and sealing all but the front SLS surface with epoxy; the electrodes were then immersed in an electrolytic cell containing either 0.2 M ferricyanide in 1 M H_2SO_4 or 0.1 M Eu^{3+} in 1 M HClO_4 .

Photocurrent action spectra (quantum yield plotted against wavelength) at room temperature are shown in Fig. 3 for two different samples of the original SLS wafer in the $\text{Eu}^{3+}/\text{HClO}_4$ electrolyte; the electrode potential was held at -0.8 V (versus SCE) and the photocurrents were cathodic. A remarkable feature is the clear presence of photocurrent peaks in the spectra; conventional GaAs photoelectrodes in the form of single crystals or thick single layers deposited by MOCVD were examined and showed the usual photocurrent onset at the band edge of GaAs followed by a smoothly increasing monotonic photocurrent that saturates to a constant value at $\sim 875 \text{ nm}$. The experimental data for the superlattice photoelectrode could be fitted extremely well by a sum of seven gaussian peaks constrained to a common

linewidth. The peak positions and heights are indicated in Fig. 3. The data could be fitted equally well using an eV scale rather than wavelength. In the former case, the fitted linewidth value was 50.6 meV .

The data show that the peak positions for photoelectrodes prepared from different regions of the SLS wafer are the same, while the peak intensities vary. We believe that the various peak intensities reflect electron transfer into the electrolyte from the discrete levels in the quantum wells. If the electrons were being injected into the electrolyte only from the lowest quantum level, then the relative peak intensities would be constant for a given quantum well geometry. Also, the observed photocurrent spectra are not consistent with that expected for absorption into the discrete levels³, followed by relaxation to and charge transfer from the lowest state. Thus, Fig. 3 shows that the SL layers are of uniform thickness, but the electron transfer probability from the discrete levels vary with position in the sample. This result is consistent with the SEM pictures (Fig. 1) showing uniform layer thicknesses but irregular outer surface layers; the outer surface in contact with the electrolyte will control the electron transfer kinetics.

Support for electron transfer from the higher lying states in the quantum well comes from the observation that the higher energy photocurrent peaks ($680 < \lambda < 750 \text{ nm}$) show markedly lower relative intensities compared with the lower-energy photocurrent peaks when a lower lying (more positive in the SCE scale) redox couple, such as ferricyanide, is used instead of Eu^{3+} . This effect is attributed to the poorer overlap of the higher lying levels in the conduction band quantum well with the energy levels in the electrolyte.

The experimental photocurrent spectra agree very well with the first six allowed ($\Delta n = 0$) transitions shown in Fig. 2. The seventh experimental peak at 1.77 eV fits to a non-allowed, $\Delta n = 1$ transition from the $n = 7$ level in the valence band to the $n = 6$ level in the conduction band; this is labelled transition $6'$ in Fig. 3. Non-allowed, $\Delta n = 1$ transitions have also been seen in SLS optical data⁶.

As seen in Fig. 3, the photocurrent response has a maximum at $\sim 750\text{--}780 \text{ nm}$, and then falls to nearly zero at wavelengths $< 680 \text{ nm}$; in this sample the response remained near zero down to 400 nm . This drop-off in photocurrent below 750 nm is, of course, highly detrimental to good solar conversion efficiencies in photoelectrochemical devices. However, we believe this drop-off is not an intrinsic characteristic of superlattice or MQW electrodes; other samples do not show this behaviour. A possible explanation is that the higher energy photons are absorbed closer

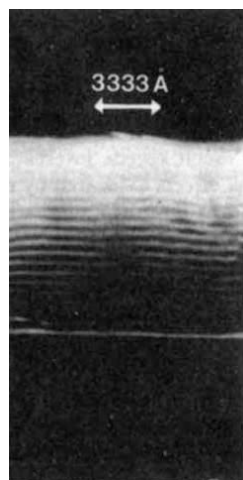


Fig. 1 SEM photograph of SLS cross-section; outer surface is at top. Wavy layers reflect strain from lattice mismatch.

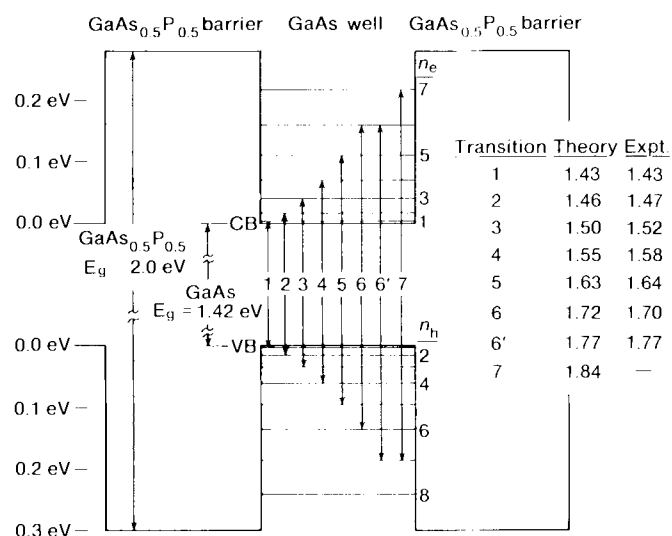


Fig. 2 Theoretical energy level scheme for $\text{GaAs}/\text{GaAs}_{0.5}\text{P}_{0.5}$ strained-layer superlattice (SLS). Allowed transitions ($\Delta n = 0$) are shown, along with a non-allowed transition ($\Delta n = 1$) from $n_h = 7$ to $n_e = 6$. The theoretical transition energies are compared with the experimental values. CB, conduction band; VB, valence band.

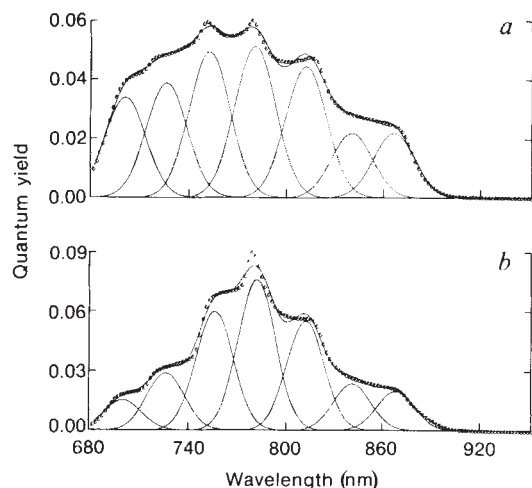


Fig. 3 Photocurrent-wavelength spectra for GaAs/GaAs_{0.5}P_{0.5} SLS. *a* and *b* represent two different regions of the SLS wafer. Solid lines, theoretical fits to seven gaussian lines with equal linewidths.

to the outer surface of the SLS, where the layer distortions and strains are greater (see Fig. 1), and hence suffer faster e^-h^+ recombination rates.

The quantum efficiencies exhibited by the present sample are consistent with electron diffusion lengths of the order of 0.05 μm . As expected, this is much lower than the values of 1–2 μm for bulk GaAs or for diffusion parallel to the superlattice layers¹². The lower values for the perpendicular direction are caused by inhibition of charge transfer by the GaAs_{0.5}P_{0.5} barriers. Significant penetration through the relatively thick barriers could be attributed to enhanced tunnelling created by strain-induced defect states in the barrier. Another possibility is that thermionic emission processes may have a role. Further work is required to understand these results more completely.

The good resolution of the photocurrent spectra even at room temperature is surprising; optical absorption and emission experiments are usually done at low temperatures to produce good spectral resolution. Further experiments as a function of SL composition, well and barrier dimensions, and temperature are in progress.

The ability of MQWs and SLs to absorb light at discrete quantum levels and permit charge transfer from these levels provides a potential mechanism for achieving the high theoretical thermodynamic conversion efficiencies predicted for hot carrier injection processes¹³. This is because carrier relaxation to the semiconductor band edges (defined for bulk material), with its attendant thermal loss, is inhibited by the presence of discrete energy levels with relatively large separation.

We thank Bruce Parkinson, Mark Spitler, and John Turner for stimulating discussions; Don Williamson of the Colorado School of Mines for computer fitting experimental data; Mowafak Al-Jassim and Kim Jones for SEM pictures; and Tim Gessert for technical assistance. This work was funded by the US Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences; J.M.O. was supported by the DOE Photovoltaic Energy Technology Division.

Received 26 February; accepted 18 April 1985.

- Esaki, L. & Chang, L. L. *Phys. Rev. Lett.* **33**, 495–498 (1974).
- Dingle, R., Gossard, A. C. & Wiegmann, W. *Phys. Rev. Lett.* **34**, 1327–1330 (1975).
- Dingle, R. *Festkörperprobleme* **15**, 21–48 (1975).
- Ploog, K. & Döhler, G. H. *Adv. Phys.* **32**, 285–359 (1983).
- Osbourne, G. C. *J. Vac. Sci. Technol.* **21**, 379–382 (1983).
- Gourley, P. L. & Biefeld, R. M. *Appl. Phys. Lett.* **45**, 749–751 (1984).
- Döhler, G. H. *Scient. Am.* **249**, 144–151 (1984).
- Brus, L. E. *J. chem. Phys.* **80**, 4403–4409 (1984).
- Rossetti, R., Ellison, J. L., Gibson, J. M. & Brus, L. E. *J. chem. Phys.* **80**, 4464–4469 (1984).
- Nozik, A. J., Williams, F., Nenadović, M. T., Rajh, T. & Mičić, O. I. *J. phys. Chem.* **89**, 397 (1985).
- Fojtik, A., Weller, H., Koch, U. & Henglein, A. *Ber. Bunsenges. Phys. Chem.* **88**, 969–977 (1984).
- Gourley, P. L., Biefeld, R. M., Zipperian, T. E. & Wiczer, J. J. *Appl. Phys. Lett.* **45**, 983–985 (1984).
- Ross, R. & Nozik, A. J. *J. appl. Phys.* **53**, 3813–3818 (1982).

Double seismic zone in Tonga

Hitoshi Kawakatsu

Department of Geophysics, Stanford University, Stanford, California 94305, USA

The clear evidence of a double seismic zone beneath Japan^{1,2} has led seismologists to search for other such zones elsewhere. Because they have been found in few other areas^{3–6}, double seismic zones are thought to be rare and to depend strongly on a peculiar combination of external conditions (such as age of the descending slab, and the convergence rate)⁷. We report here a new double seismic zone which has been found beneath the Tonga arc⁸, where the stress state in the subducting slab has been described as strongly down-dip compressional^{7,9,10}. This discovery suggests that the double seismic zone is a general feature of subducting slabs and may be found in many other subduction zones.

A double seismic zone is a zone of double-planed activity of intermediate depth earthquakes in the subducting lithosphere, usually extending from 60 to 180 km depth (Fig. 1). The two planes are generally separated by $\sim 40\text{ km}$ ^{1,2,11} at the shallow end and gradually merge at depth. The focal mechanisms of earthquakes in the upper and the lower planes tend to have compressional and tensional axes in the direction of the dip of the subducting slab, respectively¹. Double seismic zones have been reported in Tohoku, Japan^{1,2,12,13}, Kuril and Kamchatka^{3,14}, eastern Aleutians^{5,6} and Peru⁴. There have also been suggestions of double seismic zones in the central Aleutians¹⁵ and Marianas¹⁶, but the evidence is not conclusive^{7,17,18}.

Figure 2 shows a view of the general seismicity in the Tonga region with some tectonic features. The deeper portion ($>300\text{ km}$) of the slab has a high level of seismic activity and a down-dip compressional stress (or strain) state inferred from the earthquake focal mechanisms^{10,19,20}. This has led to conclusions that the slab cannot penetrate into the lower mantle in this region^{19,20}. It has also been suggested that the stress state of the intermediate depth range of the subducting slab under the Tonga arc is strongly down-dip compressional^{7,9,10}. Isacks and Molnar⁹ state that “a compressional stress is transmitted through nearly all parts of the slab”. Even after the discovery of a clear double seismic zone in Japan, neither Billington¹⁰ nor Fujita and Kanamori⁷ hinted at the possibility of a double seismic zone in the Tonga region, although they identified one event which shows down-dip tension at a depth of 186 km. However, a careful examination of the mechanisms and depths of earthquakes reveals that there are several down-dip tensional events in this region. They are located consistently deeper than and seawards of the down-dip compressional events, suggesting the presence of a double seismic zone.

Figure 3 clearly suggests the presence of a double seismic zone from a depth of $\sim 60\text{ km}$ to $\sim 200\text{ km}$. This region (azimuth 109° to 116°) was chosen because both the geometry of the subducting plate (inferred from seismicity) and the focal mechanisms suggest that the deformation of the slab occurs in a simple two-dimensional manner⁸. North of azimuth 109° ($\sim 17.5^\circ\text{S}$), the end of the subduction zone results in hinge faulting, introducing more complicated stress and strain structure in the slab. South of azimuth 116° ($\sim 24.5^\circ\text{S}$), the lateral bend of the slab near 26°S prevents a simple two-dimensional view of the subduction process.

Cross-sections with narrower azimuth ranges (every 1° along the strike of the trench)⁸ including smaller events show that the down-dip compressional and tensional events are consistently located on the upper and the lower edge of the zone of seismicity, respectively, in every section. The result of relative relocation (Fig. 4) supports the presence of a double seismic zone. The down-dip tensional events in Figs 3 and 4 are consistently located seaward of and below the down-dip compressional events, exactly as expected for a double seismic zone.

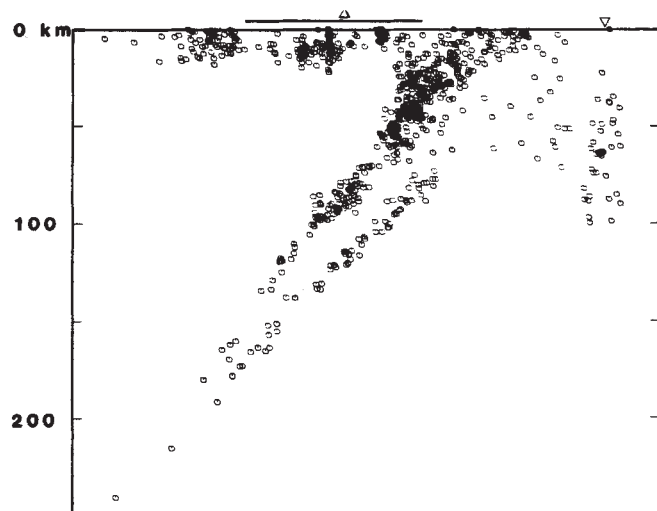


Fig. 1 Double seismic zone beneath the northern Honshu arc, Japan (after ref. 1). The precise determination of the microearthquake hypocentres using data from the seismic network of Tohoku University revealed a clear separation of two seismic zones. The figure is vertically exaggerated by a 2:1 ratio.

What are the implications of this discovery of the double seismic zone beneath the Tonga arc? First, the double seismic zone is a rather persistent feature of subducting slabs. Tonga has the most active deep seismicity but the compressional strain (due either to inability of the slab to penetrate into the lower mantle^{19,20} or to the strong viscous resistance to the subduction in the mesosphere⁹) transmitted from below 200 km depth cannot overprint and erase the pre-existing double seismic zone. On this basis, I predict that there is also a double seismic zone beneath the Izu-Bonin arc, even though it has active deep seismicity and down-dip compressional stresses. Furthermore, I speculate that a double seismic zone will be observed in most subduction zones (at least in those areas where both down-dip compressional and tensional events are already observed^{7,9}). The observation of double seismic zones has been prevented by the lack of high-quality local networks in most subduction zones and the fact that the time span of reliable observations of earthquake focal mechanisms is too short. It has also been difficult to distinguish shallow (<80 km) down-dip tensional events from shallow-angle thrust faulting events, which are very common in subduction zones, without detailed wave-form analysis. This is because they have similar focal mechanisms and the reported depths (obtained by the International Seismology Centre (ISC) or by preliminary determination of epicentre) of shallow events are not reliable. A systematic analysis of these events, which can now be undertaken through the Global Digital Seismograph Network²¹⁻²³, will help to identify more down-dip tensional events.

Second, this discovery constrains possible mechanisms for the origin of the double seismic zone. It excludes geometrical sagging^{24,25} as a possible mechanism since the strong down-dip compressional strain transmitted from below 200 km could easily overprint strain features caused solely by sagging. The simple geometrical unbending^{15,16,26} with a brittle-plastic rheology suggested by Tsukahara²⁶ seems to be the best description of the physical mechanism of the double seismic zone.

Above a depth of ~60 km, oceanic lithosphere bends downwards when it subducts into the mantle. The fact that the subducting lithosphere appears to be straight in the mantle below a depth of ~200 km indicates that the lithosphere must unbend in a depth range from ~60 to ~200 km to become straight below a depth of ~200 km. A simple geometrical argument shows that the strain rate due to this unbending is of the order of 10^{-15} to 10^{-14} s^{-1} above a depth of ~100 km where most of unbending is taking place and is high enough to account for all the energy released by earthquakes of the double seismic zone⁸. Even below a depth of ~100 km where the slab appears to be straight, the

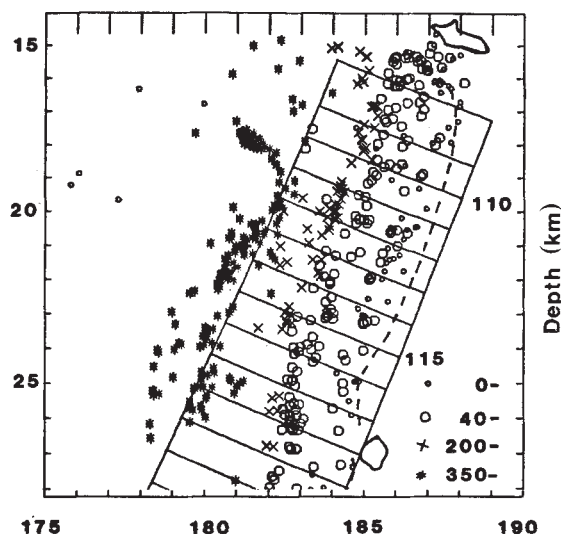


Fig. 2 Schematic view of the general seismicity of the Tonga subduction zone. The depth-constrained ISC events are plotted with tectonic features. The broken line represents the trench. The lines perpendicular to the trench are parts of the great circles which pass through the pole of the arc and the numbers denote the azimuth of the great circle measured clockwise from the north at the pole of the arc. The pole of the arc is located at 20.6° N, 102.8° E, so that the 100-km isodepth contour of Isacks and Barazangi⁴ is 90° distant away from the pole. The two lines parallel to the strike of the trench are at the distances 87.5° and 92.5°.

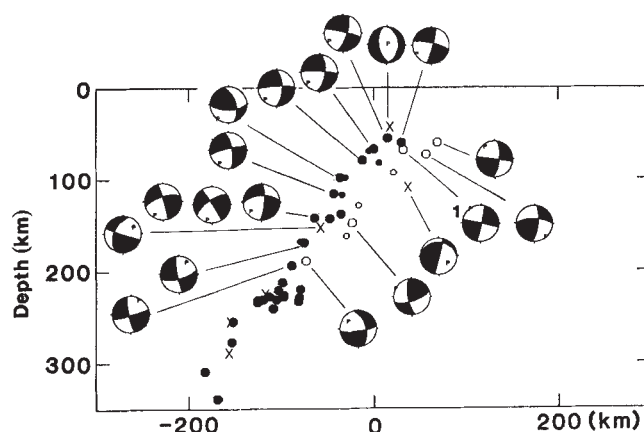


Fig. 3 Cross-section perpendicular to the trench. All the intermediate-depth (60 to 350 km) events occurring between azimuths 109° and 116° for which focal mechanisms are available, are plotted. ●, Down-dip compressional; ○, down-dip tensional events; ×, events which are not either compressional or tensional. The small circles are events for which only the type of the mechanism can be determined⁸. The depths are all constrained by p-P times reported with ISC. For the down-dip tensional events, depths are also checked by WWSSN seismograms. For events shallower than 200 km, the near side of the focal sphere is also plotted. Event 1: the depth (61 km)²³ of this large ($M_0 = 1.4 \times 10^{28} \text{ dyn cm}$)²³ normal fault event (22 June, 1977) is controversial²³. The aftershock distribution³⁰ and the excitation of overtone surface waves (E. Okal, personal communication) seem to suggest that the faulting extended deeper than 100 km. Thus, the depth given here is probably the shallowest estimate.

unbending strain rate can easily be of the order of 10^{-17} to 10^{-16} s^{-1} (an unbending strain rate of 10^{-17} s^{-1} of a 100-km thick slab, subducting 10 cm yr^{-1} , corresponds to a change of the dip angle of $\sim 0.05^\circ/100 \text{ km}$, which is impossible to detect from seismicity⁸). On the other hand, the strain rate due to thermo-elasticity^{27,28} (10^{-18} to 10^{-17} s^{-1})²⁸ is too low to be the main cause of the double seismic zone. Although all the other proposed mechanisms may contribute to its occurrence, the double seismic zone can be satisfactorily explained by the simple

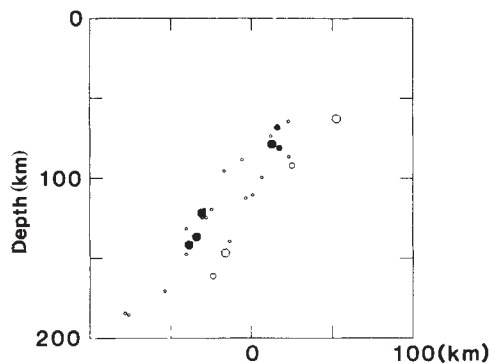


Fig. 4 To assign appropriate relative errors to the hypocentral parameters, a relative relocation was performed⁸. The method is similar to one used by Samowitz and Forsyth¹⁶, except that here only depth constrained events are used. Data are selected from the ISC bulletin from 1971 to 1979 with the following conditions; pP-P constrained depths between 60 and 200 km; >30 P arrival observations. For the first iteration, all events are relocated with the depths fixed at the pP-P depths using the P-wave arrival times reported at stations between 30° and 100° distance and ISC locations as the starting guesses; 440 stations were used. After the first iteration, the stations which have arrivals from >10 earthquakes are selected and assigned station adjustments as a linear function of distance between stations and earthquakes by the least squares method. This leaves 113 stations. For the second iteration, travel times are first adjusted by using the station adjustments estimated in the first iteration. Each event was then relocated separately using only the arrivals from those 113 stations. The estimated relative errors (90% confidence intervals) are usually smaller than the largest circles. Symbols are as in Fig. 3. The smallest circles denote events for which no information about their focal mechanism is known. Although a clear separation of two seismic zones seems to exist, it could be coincidence¹⁷. Therefore, what is most important is that the down-dip compressional events are still always located above the down-dip tensional type events even after the relocation.

model of geometrical unbending. The cause of this unbending is uncertain and probably varies from place to place. For example, pressure due to mantle flow²⁹ (or equivalently, dynamical sagging²⁵) or the weight of the slab itself can unbend the slab.

For a given magnitude of the strain rate, the distance between the two seismic zones depends strongly on the rheological properties and the thermal structure of the descending lithosphere²⁶. Thus, as pointed out by Tsukahara²⁶, studies of the fine structure of double seismic zones in many areas will give important new information on the thermal and rheological properties of the subducting lithosphere⁸.

I thank Bob Geller, Eric Peterson, Rick Schult and Norm Sleep for reviewing the manuscript. Comments by Drs Klaus Jacob and Bob Engdahl also improved the manuscript. This work was supported by NSF grants EAR82-18961 and EAR83-06555.

Received 24 January; accepted 16 April 1985.

- Hasegawa, A., Umino, N. & Takagi, A. *Tectonophysics* **47**, 43–58 (1978).
- Hasegawa, A., Umino, N. & Takagi, A. *Geophys. J.R. astr. Soc.* **54**, 281–296 (1978).
- Veith, K. F. thesis, Southern Methodist Univ., (1978).
- Isacks, B. L. & Barazangi, M. *Island Arcs, Deep Sea Trenches and Back-Arc Basins* Vol. 1, 99–114 (American Geophysical Union, Washington DC, 1977).
- Reyners, M. & Coles, K. *J. geophys. Res.* **87**, 356–366 (1982).
- House, L. S. & Jacob, K. H. *J. geophys. Res.* **88**, 9347–9373 (1983).
- Fujita, K. & Kanamori, H. *Geophys. J. R. astr. Soc.* **66**, 131–156 (1981).
- Kawakatsu, H. thesis, Stanford Univ. (1985).
- Isacks, B. L. & Molnar, P. *Nature* **223**, 1121–1124 (1969); *Rev. Geophys. Space Phys.* **9**, 103–174 (1971).
- Billington, S. thesis, Cornell Univ. (1980).
- Stefani, J. P., Geller, R. J. & Kroeger, G. C. *J. geophys. Res.* **87**, 323–328 (1982).
- Yoshii, T. *Tectonophysics* **55**, 349–360 (1979).
- Kawakatsu, H. & Seno, T. *J. geophys. Res.* **88**, 4215–4230 (1983).
- Suzuki, S., Sasatani, T. & Motoya, Y. *Tectonophysics* **96**, 59–76 (1983).
- Engdahl, E. R. & Scholz, C. H. *Geophys. Res. Lett.* **4**, 473–476 (1977).
- Samowitz, I. & Forsyth, D. *J. geophys. Res.* **86**, 7013–7021 (1981).
- Engdahl, E. R. & Fijita, K. *J. geophys. Res.* **86**, 7023–7024 (1981).
- Topper, R. E. thesis, Univ. Colorado (1978).
- Richter, F. M. *J. geophys. Res.* **84**, 6783–6795 (1979).
- Giardini, D. & Woodhouse, J. H. *Nature* **307**, 505–509 (1984).
- Dziewonski, A. M., Chou, T. A. & Woodhouse, J. H. *J. geophys. Res.* **86**, 2825–2852 (1981).
- Dziewonski, A. M. & Woodhouse, J. H. *J. geophys. Res.* **88**, 3247–3271 (1983).

- Giardini, D. *Geophys. J.R. astr. Soc.* **77**, 883–914 (1984).
- Yoshii, T. *Kagaku* **47**, 170–176 (1977).
- Sleep, N. H. *J. geophys. Res.* **84**, 4565–4571 (1979).
- Tsukahara, H. *J. phys. Earth* **28**, 1–15 (1980).
- Goto, K. & Hamaguchi, H. *Prog. Abstr. seism. Soc. Jap.* **2**, 36 (1978).
- House, L. S. & Jacob, K. H. *Nature* **295**, 587–589 (1982).
- Yokokura, T. *Tectonophysics* **77**, 35–62 (1981).
- Silver, P. G. & Jordan, T. H. *J. geophys. Res.* **88**, 3273–3293 (1979).

The influence of subduction processes on the geochemistry of Japanese alkaline basalts

Eizo Nakamura*, Ian H. Campbell*† & Shen-su Sun†‡

* Department of Geology, University of Toronto, Toronto, Ontario, Canada M5S 1A1

† Research School of Earth Sciences, The Australian National University, GPO Box 4, Canberra, ACT 2601, Australia

‡ Division of Petrology and Geochemistry, Bureau of Mineral Resources, Geology and Geophysics, Canberra, ACT 2601, Australia

The composition of volcanic rocks erupted in complex plate tectonic settings can provide information on the nature of the underlying mantle. We show here that the geochemistry of alkali basalts from Japan and eastern Asia varies systematically with distance from the Japanese island-arc. Samples from northeastern Japan, relatively close to the Japan Trench, are enriched in K, Sr, Ba and Rb and depleted in Ta, Nb and Ti as compared with samples from southwestern Japan. Both sets show an island-arc influence on their composition, but alkali basalts from still further west (Korea and northeastern China) have chemistries which are indistinguishable from ocean island basalts. We suggest that the north-eastern island-arc type of alkali basalts were derived from a 'normal' upper mantle source altered by fluids or melts released from the underlying subducted Pacific plate. The extent of this island-arc-related alteration decreases with distance from the trench.

Major and selected trace-element concentrations of 200 Tertiary to Recent alkali basalts were determined by a combination of X-ray fluorescence (XRF) and instrumental neutron activation analysis (INAA) using the method described by Barnes and Gorton¹. Most of the samples analysed are nepheline normative and have the following characteristics: SiO₂ contents of less than 52 wt%, Ni contents of greater than 100 p.p.m., Cr contents greater than 150 p.p.m. and no detectable Eu anomalies. Samples with these characteristics are assumed to have retained their primitive chemistry. Olivine ± clinopyroxene fractionation may have produced a slight increase in the concentration of the incompatible elements in some samples but any change in ratios of these elements will be small.

Selected analyses are plotted on normalized abundance pattern (NAP) diagrams^{2–5} (Fig. 1). These diagrams give the concentration of selected incompatible elements normalized to their concentration in the primitive mantle. The elements are arranged systematically in order of increasing incompatibility, with the most incompatible elements plotted on the left. Figure 1a gives NAPs for an average depleted mid-ocean ridge basalt^{3,4}, typical Hawaiian tholeiite⁶ and alkali basalt⁶, an average continental alkali basalt from Kenya⁷, an average island-arc calc-alkaline basalt⁴ and an average arc tholeiite⁴. Both island-arc rock types are characteristically enriched in K, Ba, Sr and Rb and depleted in Ta, Nb and Ti (weakly) relative to adjacent elements in the NAP diagram.

NAP diagrams for samples from Rishiri Island (northeastern Japan), southwestern Japan, Jejudo Island (Korea) and Datong (northeastern China) are plotted in Fig. 1b, c and d respectively. They can be divided into three types: an island-arc type enriched in K, Ba, Sr and Rb and depleted in Ta or Nb and Ti, a transitional type which shows a much weaker island-arc signature, and a normal type with a NAP typical of ocean island and continental alkali basalts. The island-arc type is confined

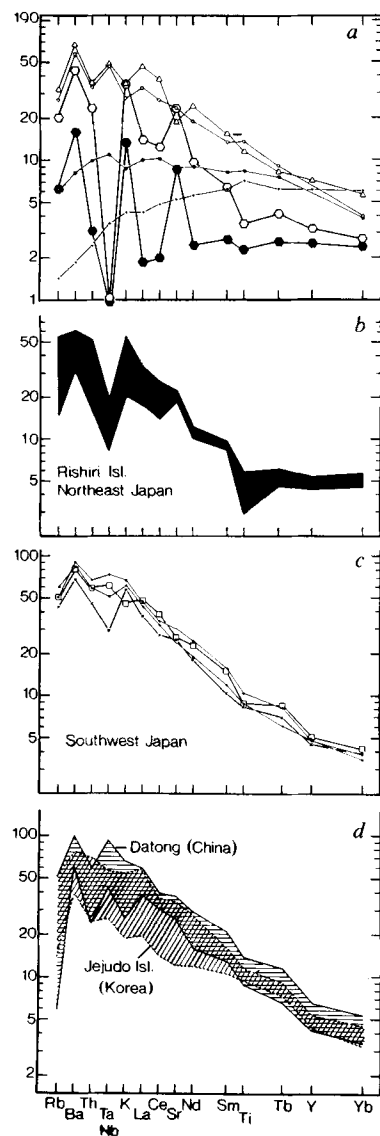


Fig. 1 *a*, Normalized abundance pattern (NAP) diagram for a typical ocean island alkali basalt⁶ (○) and tholeiite⁶ (●) from Hawaii, average continental alkali basalt (△) from Kenya⁷, average island-arc calc-alkali basalt⁴ (○) and tholeiite⁴ (●) and average depleted mid-ocean ridge basalt⁴ (+). *b*, NAP diagram showing the range of value for six alkali basalts from Rishiri Island, northeastern Japan. The patterns show a strong island-arc signature. *c*, Selected NAPs for transitional alkali basalts from southwestern Japan. The patterns were selected to show the highest K^* value [1.51 (×)], the lowest value [0.84 (□)] and two intermediate values (●). *d*, Variations in the NAPs for normal alkali basalts from Datong, northeastern China and Jeju Island, Korea. Note all element abundances in *a*–*d* are normalized to their concentration in the primitive mantle.

to northeastern Japan, the transitional type to southwestern Japan and the normal type to Korea and northeastern China.

The geographical distribution of the three types of alkali basalt is summarized in Fig. 2 where the average value of $K^* = 2[K/(Ta + La)]_n$ (abundances of all three elements are normalized to primitive mantle values) is plotted for the alkali volcanic centres studied in this project and for primary magma compositions of sub-alkali volcanic centres studied by Sakuyama and Nesbitt⁸. This ratio was chosen because Ta and La are adjacent to K on the NAP diagram. K^* can be taken as a measure of the amount of K enrichment and therefore the intensity of the island-arc signature. As shown in Fig. 1*a*, ocean island and continental alkali basalts are typically depleted in K relative to La and Ta (that is, $K^* < 1$).

K^* values for primitive of sub-alkali basalts from northeastern Japan are greater than 2.8, with highest values occurring closest

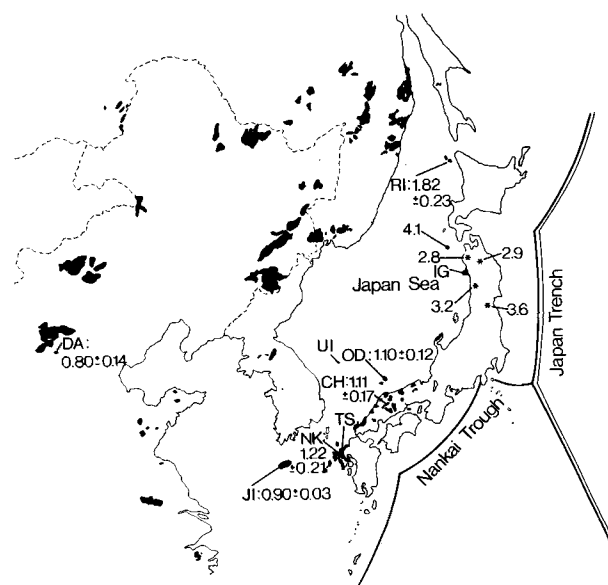


Fig. 2 A map of the Japan Sea area showing distribution of Cenozoic alkali basalts and $K^* = 2 \times [K/(Ta + La)]_n$ values for each centre, with standard deviation for the alkali volcanic centres studied. Additional data from Sakuyama and Nesbitt⁸ for one alkali (●) and several tholeiitic and calc-alkaline volcanic centres (*) in northeastern Japan are also plotted. Note that northeastern Chinese Cenozoic basalts³⁶ have K^* values < 1 (negative K anomalies). Symbols: RI, Rishiri Island; IG, Ichinomegata; OD, Oki-Dogo; CH, Chugoku; NK, northwestern Kyushu; TS, Takashima; JI, Jeju Island; UI, Ulungdo Island; DA, Datong.

to the Japan Trench. The highest K^* value for alkali basalts from northeastern Japan (4.1) is also from the centre closest to the trench, with the value for Rishiri Island, further from the trench in this area, being appreciably less. The average value of K^* for centres from southwestern Japan varies between 1.1 and 1.2, although individual samples may have values as low as 0.84. Samples from Korea and northeastern China invariably have values below 1.0. Figure 2 shows that K^* values for alkali basalts decrease away from the Pacific plate trench. Variations in Ba, Sr and Rb abundances show similar trends to the K^* values. The concentrations of major elements and all other trace elements, including Zr, Y and rare earth elements (REE), are similar for all types of alkali basalts in the area studied.

The observation of decreasing K^* values away from the Pacific plate trench is apparently at odds with Kuno's study⁹ which showed that that K_2O content of Japanese volcanic rocks increase with distance from the trench. We interpret the increase in K_2O to be due to decreasing degrees of partial melting and the change in K^* values to be due to a decrease in subduction-related alteration of the mantle with increasing distance from the trench. Thus the two observations are not incompatible.

The lack of a significant difference in major and trace elements (Th, Zr, Y and REE) for samples from Japan, Korea and northeastern China suggests that the alkali basalts from each of these areas formed by similar processes. If this is correct, the observed enrichments in Ba, K, Sr and Rb must reflect differences in the concentration of these elements in the source. Since the Japanese alkali basalts are found above the subducted Pacific plate except in the northwestern Kyushu area, the obvious explanation for the differences is that the mantle above the subduction zone has been contaminated by K, Ba, Sr and Rb bearing metasomatic solutions and/or partial melts derived from the slab^{10,11}.

Alkali basalts from southwestern Japan show little evidence of an island-arc signature in spite of their proximity to the Philippine Sea plate subduction zone. This is a young subduction zone and the angle of subduction is shallow^{12–14}. We believe that the island-arc type geochemical signature is absent from

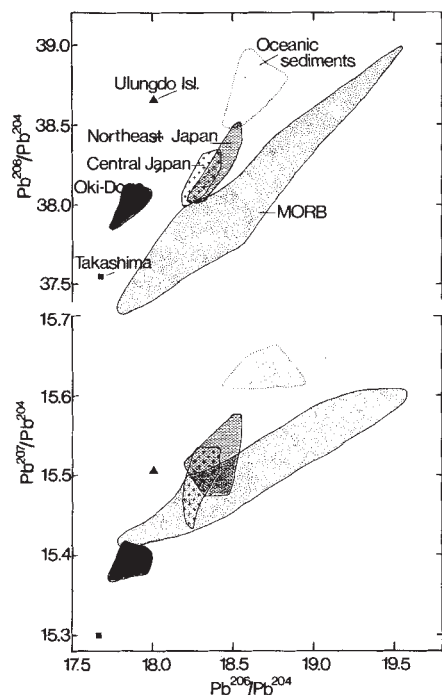


Fig. 3 Lead isotope compositions of Japanese alkaline, tholeiitic and calc-alkaline basalts¹⁸⁻²⁰, oceanic sediments^{3,37,38} and MORB^{39,40}. All those analyses carried out by the PbS method were recalculated to 'absolute values' for comparison with those analysed by the silica-gel method using a correction factor of -0.35% per mass unit. The data for Ulungdo Island are unpublished analyses by D. Nelson (personal communication). Pb isotopic compositions of basalts in northeastern and central Japan show a mixing relationship between oceanic sediments and MORB, and become less radiogenic with increasing distance from the Japan Trench. In central Japan, the compositional variation with distance is less regular because of the small difference in the distance between volcanic centres. However, the volcanic centre nearest to the trench, Ohshima Island, has the most radiogenic Pb composition. Pb isotope compositions of alkali basalts from southwestern Japan are similar to basalts from Hawaii. In spite of having similar $^{206}\text{Pb}/^{204}\text{Pb}$ ratios to MORB, their $^{208}\text{Pb}/^{204}\text{Pb}$ ratios are much higher than MORB, suggesting an evolutionary history with a higher Th/U ratio and a more primitive source.

this area because the contaminated mantle above the Philippine Sea plate has not reached sufficient depth to be involved in the production of alkali basalts.

Our conclusion that the alkali basalts of northeastern and southeastern Japan have come from mantle sources which have undergone different degrees of subduction-related alteration is supported by studies of mantle xenoliths. Xenoliths from Ichinomegata, northeastern Japan, have abundant hydrous minerals¹⁵ and contain water-rich fluid inclusions^{16,17}, whereas those from southwestern Japan contain a high-temperature, nearly anhydrous mineral assemblage¹⁵. These observations are consistent with the hypothesis that the mantle source of Japanese alkali basalts is strongly contaminated in the north-east, but only weakly contaminated in the south-east.

Further support for this interpretation is provided by available Pb¹⁸⁻²⁰, Sr^{21,22} and Nd²² isotopic studies. Basalts from northeastern and central Japan have a strong component of Pb from oceanic sediments (Fig. 3) which decreases with distance from the trench, that is, the sedimentary Pb component is the strongest in arc tholeiites and weakest in alkali basalts. In contrast, alkali basalts from southwestern Japan show Pb systematics similar to basalts from nonradiogenic ocean islands such as Hawaiian tholeiites²³. These observations suggest that the fluids (or melts) which contaminated the mantle in northeastern Japan had a strong sedimentary component and that this component decreased in importance with increasing distance from the Pacific plate trench. The implication is that the sedimentary components which contaminate the overlying wedge are con-

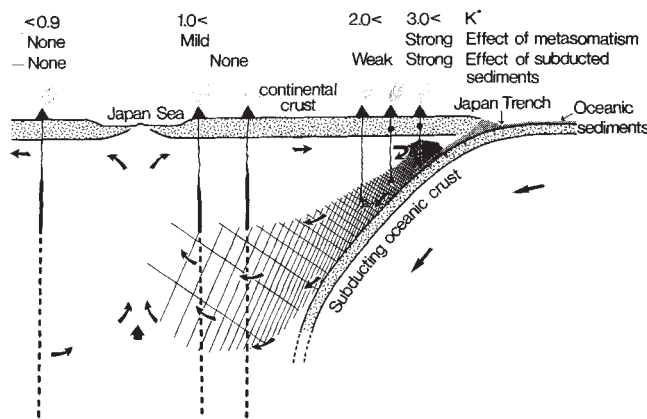


Fig. 4 A diagrammatic model of magma generation above the Japanese subduction zone. The subducted slab dehydrates or melts as it descends, releasing a fluid which contaminates the overlying mantle (cross-hatched zone). The intensity of this contamination decreases with increasing distance from the trench. Oceanic sediments (dotted zone) are also carried down by the subducted slab, but dehydrate (or melt) at shallow depths so that their influence is confined to volcanic centres close to the trench. All volcanic rock types from northeastern Japan have come from a strongly contaminated MORB-type source. The alkali basalts of southwestern Japan are believed to originate deeper in the mantle from an OIB-type source and to have interacted chemically with an overlying contaminated MORB-type mantle during their ascent. Alkali basalts from China and Korea were probably produced by a similar process but did not interact with a subduction-contaminated mantle.

sumed at relatively shallow depths. Sr and Nd isotopic studies of basalts from northeastern Japan also require the mantle to be contaminated by a component derived from subducted oceanic sediments^{4,22}.

Figure 4 is a summary diagram which relates our geochemical data and the available isotopic data for Japanese volcanics to existing petrological²⁴⁻²⁹ and convection models³⁰⁻³² for island-arc systems. Pb, Sr and Nd isotopic studies require at least two types of mantle to produce the volcanic rocks of Japan; a mid-ocean ridge (MORB) type and an ocean island (OIB) type. The tholeiitic and alkalic volcanics of northeastern Japan required a MORB-type source which has been strongly contaminated by fluids and/or melts derived from the subducted Pacific plate. The intensity of this contamination decreases with distance from the trench. We suggest that the plumes that give rise to the alkali volcanics of southwestern Japan originate from an OIB-type source at a much greater depth. As the ascending plumes rise, they assimilate some of the overlying contaminated mantle and, in doing so, acquire a weak island-arc geochemical signature. The alkali basalts of northeastern China and Korea show no evidence of an island-arc signature and are geochemically indistinguishable from ocean island alkali basalts. Chemical and isotopic variations observed in these alkali basalts may have resulted from interaction between heterogeneous plume source material and the continental lithosphere.

The absence of a sedimentary Pb component from alkali basalts of southwestern Japan, which show a mild island-arc geochemical signature, is difficult to reconcile with a geochemical model which requires oceanic sediments to be subducted in an island-arc system and recycled through the mantle^{33,34}. Our observations suggest that the sedimentary component is consumed at shallow depths during subduction through repeated dehydration and partial melting as the oceanic lithosphere sinks into the mantle. Sedimentary components may be recycled into the mantle as a consequence of subduction related to continental collision³⁵, recycling of the mantle wedge above the subduction zone or delamination and recycling of the lithospheric mantle³⁵ previously processed through the island-arc environment.

We thank E. Takahashi, H. Takamura, T. Yoshida, M.-W. Lee, Y.-K. Kim and C.-Y. Cui for donating samples from Oki-Dogo,

Chugoku, Rishiri Island, Jejudo Island, Ulungdo Island and Datong respectively. S.-S. publishes with permission of the Director of Bureau of Mineral Resources Australia. We also thank A. W. Hofmann for his critical comments.

Received 29 November 1984; accepted 11 April 1985.

1. Barnes, S.-J. & Gorton, M. P. *Geostand. Newslett.* **8**, 17-23 (1984).
2. Sun, S. S. *Geochim. cosmochim. Acta* **46**, 179-192 (1982).
3. Sun, S. S. *Phil. Trans. R. Soc. A297*, 409-446 (1980).
4. Sun, S. S., Nesbitt, R. W. & Sharaskin, A. Y. *Earth planet. Sci. Lett.* **44**, 119-138 (1979).
5. Thompson, R. N., Scott, J. *Geol.* **18**, 50-107 (1982).
6. Chen, C. Y. & Frey, F. A. *Nature* **302**, 785-789 (1983).
7. Baker, B. H., Gole, G. G., Leeman, W. P. & Lindstrom, M. M. *Contr. Miner. Petrol.* **64**, 303-332 (1977).
8. Sakuyama, M. & Nesbitt, R. W. *Proc. ISEF Field Conf. Open Magmatic Systems*, Taos, New Mexico, 138-140 (Southern Methodist University, 1984).
9. Kuno, H. *Bull. Volcanol.* **29**, 195-222 (1966).
10. Pearce, J. A. in *Continental Basalts and Mantle Xenoliths* (eds Hawkesworth, C. J. & Norry, M. J.) 230-249 (Shiva, Orpington, 1983).
11. Loughnan, F. C. *Chemical Weathering of Silicate Minerals* (Elsevier, New York, 1969).
12. Nakanishi, I. *Tectonophysics* **69**, 1-35 (1980).
13. Sugimura, A. *J. phys. Earth* **20**, 147-164 (1972).
14. Uyeda, S. *Tectonophysics* **81**, 133-159 (1982).
15. Takahashi, E. *Bull. Volcanol.* **41**(4), 529-547 (1978).
16. Roedder, E. *Am. Miner.* **50**, 1746-1782 (1965).
17. Trial, A. F., Rudnick, R. L., Ashwal, L. D., Henry, D. J. & Bergman, S. C. *EOS* **65**, 306 (1984).
18. Kurasawa, H. *Geochem. J.* **2**, 11-28 (1968).
19. Hedge, C. E. & Knight, R. J. *Geochem. J.* **3**, 15-24 (1969).
20. Tatsumoto, M. & Knight, R. I. *Geochem. J.* **3**, 53-86 (1969).
21. Notsu, K. *J. Volcan. geotherm. Res.* **18**, 531-548 (1983).
22. Nohda, S. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **52**, 264-276 (1981).
23. Tatsumoto, M. *Earth planet. Sci. Lett.* **38**, 63-87 (1978).
24. Ringwood, A. E. *J. Geol. Soc. Lond.* **130**, 183-204 (1974).
25. Nicholls, I. A. & Ringwood, A. E. *J. Geol.* **81**, 285-300 (1973).
26. Fyfe, W. & McBirney, A. *Am. J. Sci.* **275A**, 285-297 (1975).
27. Sekine, T. & Wyllie, P. J. *Contr. Miner. Petrol.* **79**, 368-374 (1982).
28. Sekine, T. & Wyllie, P. J. *J. Geol.* **91**, 511-528 (1983).
29. Wyllie, P. J. & Sekine, T. *Contr. Miner. Petrol.* **79**, 375-380 (1982).
30. Ida, Y. *J. geophys. Res.* **88**, 7449-7456 (1983).
31. Toksoz, M. N. & Hsui, A. T. *Tectonophysics* **50**, 177-196 (1981).
32. Jurdy, D. M. & Stefanick, M. *Tectonophysics* **99**, 191-206 (1983).
33. Armstrong, R. L. *Rev. Geophys.* **6**, 175-199 (1968).
34. White, W. M. & Hofmann, A. W. *Nature* **296**, 821-825 (1982).
35. Thompson, R. N., Morrison, M. A., Hendry, G. L. & Party, S. J. *Phil. Trans. R. Soc. A310*, 549-590 (1984).
36. Zhou, X. & Armstrong, R. L. *Earth planet. Sci. Lett.* **58**, 301-329 (1982).
37. Chow, T. J. & Patterson, C. C. *Geochim. cosmochim. Acta* **26**, 263-308 (1962).
38. Chow, T. J. in *Recent Research in the Field of Hydrosphere, Atmosphere and Nuclear Geochemistry*, 179-183 (Maruzen, Tokyo, 1964).
39. Dupré, B. & Allégre, C. J. *Nature* **286**, 17-22 (1980).
40. Cohen, R. S. & O'Nions, R. K. *J. Petrol.* **23**, 299-324 (1982).

Microzones surrounding phytoplankton form the basis for a stratified marine microbial ecosystem

James G. Mitchell, Akira Okubo & Jed A. Fuhrman

Marine Sciences Research Center, State University of New York, Stony Brook, New York 11794-5000, USA

Although there are 10^7 – 10^9 bacteria per litre of sea water and up to 50% of marine primary productivity passes through them¹, no direct information exists on small-scale spatial distributions of individual bacteria. It has been proposed that bacteria cluster around particulate nutrient sources^{2,4}, but evidence for this is indirect, based on the presence of chemotaxis in isolates and a broad range— 10^{-9} – 10^{-4} M—of uptake affinity constants³. These data have led to the suggestion that there exist oceanic microzones having utilizable dissolved organic carbon (uDOC) orders of magnitude above background concentrations⁴. Here, our calculations of size (10^{-1} cm), utilizable DOC concentrations (10^{-9} – 10^{-7} M) and lifetimes of these nutrient gradients suggest that bacteria cluster around phytoplankton in low-turbulence layers and sheets comprising the thermocline. There, phytoplankton may sink slowly enough for bacteria to keep up with them, and mixing times are long enough to allow nutrient gradients to develop and bacteria to track these gradients. However, very favourable conditions are required for microzone concentrations to be significantly above background.

The size of the nutrient field was calculated from the excretion rate per cell and the dispersion of nutrients away from the cell.

We assumed that 40% of a phytoplankton's primary production is excreted; this represents a generous intermediate value from the wide range cited in the literature^{5,6}. We used high values for phytoplankton productivity⁷ (10^{-4} g C per l per day, over a 12-light period) and cell concentration⁷ (10^6 cells per l) (a more typical value is 10^{-5} g C per l per day). Based on recalculations of Chrost and Faust's⁸ data, we used 100 g as the weight of carbon per mole, and this indicates that 85% of the molecules that phytoplankton excrete have a relative molecular mass <500; amino acids and sugars are the major components of this size fraction⁹. These values result in a total flux out of a cell of 10^{-17} mol s⁻¹.

From the total nutrient flux per cell and the diffusivity of nutrient away from the cell, we determined the dimension of the microzone around a phytoplankton using equation (1)¹⁰

$$\text{Radius} = \frac{(\text{total flux})}{(4\pi)(\text{concentration})(\text{diffusivity})} \quad (1)$$

Here, radius means the distance at which the nutrient gradient emanating from the cell is enhanced 10% above the background concentration. We used a background concentration of 10^{-8} M or 10^{-11} mol cm⁻³ (ref. 11), a relatively low value that again favours microzones (a more typical value, including carbohydrates, is 10^{-7} M; ref. 12). Consequently, 10% of the background level is 10^{-9} M (represented by concentration in equation (1)). Equation (1), using 10^{-5} cm² s⁻¹ for diffusivity¹³, gives a radius of 8×10^{-2} cm or ~ 1 mm. At this distance from the cell the concentration is theoretically 1.1×10^{-8} M, compared with 1×10^{-8} M for background. For the more typical values cited above, the radius is $<10^{-3}$ cm.

To determine the significance of the larger microzones, we are interested in the complete range of concentrations. The concentration at the phytoplankton surface represents the maximum enhanced concentration. To estimate this maximum surface concentration, equation (1) was rearranged, and we assumed a relatively small diatom radius of 5×10^{-4} cm to minimize surface area and thus maximize surface concentration (a more typical value is about three times as large). The resulting maximum concentration, under optimum conditions, is $\sim 10^{-7}$ M. For the more typical values cited above, this maximum is $\sim 5 \times 10^{-9}$ M, or barely above background. Thus, if much higher concentrations exist in the literature, diffusion must be restricted by a physical barrier, such as mucus.

One must also compare the formation time of the zone with its rate of disintegration by turbulence. Based on Carslaw and Jaeger¹⁰, we calculate that it takes at least 10^2 s for the leading edge of the excreted material to reach a radius of 1 mm. Table 1 demonstrates that nutrients are dispersed long before reaching a radius of 1 mm, except in low-turbulence conditions such as

Table 1 Lifetimes of Kolmogorov eddies within the range of turbulence found in the ocean

Energy (E) dissipation (cm ² s ⁻³)	Eddy			
	Length (cm)	Velocity (cm s ⁻¹)	Lifetime (s)	Typical location
10^{-2}	0.10	0.10	1.0	Mixed layer
10^{-3}	0.18	0.056	3.2	Mixed layer
10^{-4}	0.32	0.030	10	
10^{-5}	0.56	0.018	32	
10^{-6}	1.0	0.010	100	Thermocline
10^{-7}	3.2	0.0056	320	Thermocline

Kolmogorov eddies are the hydrodynamic 'units' which contain a phytoplankton and its microzone. The necessary equations are from Batchelor²⁷. Values are from Ozmidov²⁸ and from Gregg²⁹. Length = $(v^3/E)^{1/4}$; velocity = $(vE)^{1/4}$; lifetime = $(v/E)^{1/2}$, where v = kinematic viscosity of water (10^{-2} cm² s⁻¹).

the thermocline. The thermocline is composed of 1–5-m-thick 'layers', regions of minimal turbulence separated by 1–10-cm-thick sheets of near-laminar flow¹⁴. Sheets and layers also characterize the 'mixed' layer on calm days¹⁴. We suggest that sheets and layers are the most likely location for microzones for two reasons: (1) in these areas a nutrient gradient has time to reach steady state around a phytoplankter, and (2) bacteria have time to track this gradient chemotactically. The latter point is explained below, but essentially bacteria cannot move by chemotaxis fast enough to form a cluster in less than ~100 s.

For bacteria to cluster around a phytoplankter, they must find, enter and stay within the nutrient gradient. Bacterial sensitivity to chemical gradients^{15,16} and their ability to orient to them is well documented^{17,18}. However, in conventional chemotaxis studies, the nutrient source is stationary and the bacteria are given minutes to hours to orient^{2,19}. Motile bacteria swim, stop, randomly reorient, and swim again. Bacteria approach a nutrient source because runs between stops are longer when they move up a gradient than when they move down one¹⁷. Random reorientation means that bacteria move discontinuously up a gradient in a biased random walk. Consequently, net movement in a gradient is only ~1–3 $\mu\text{m s}^{-1}$, about an order of magnitude slower than their average swimming speed of 10–30 $\mu\text{m s}^{-1}$ (ref. 20). These values derive from laboratory studies of enteric bacteria. The rates for marine bacteria are probably similar, but unknown; if they differ significantly our conclusions are altered accordingly.

It is important to compare the rate of bacterial chemotaxis with the motion of phytoplankton. In the ocean's mixed layer an average non-motile phytoplankter, such as a diatom, sinks at a rate of ~10 $\mu\text{m s}^{-1}$ (1 m per day; ref. 21) or about three times the speed at which a bacterium can move chemotactically towards a nutrient source. This suggests that bacteria have difficulty in keeping up with most diatoms and probably cannot overtake or maintain themselves within the nutrient gradient, making clustering of bacteria to form microzones unlikely in the mixed layer. We chose sinking as an example of cell movement because the average values are well documented²¹. But any movement of the nutrient source (relative to the immediately surrounding water) that is more rapid than the bacterial chemotactic rate (~3 $\mu\text{m s}^{-1}$) prevents clustering. Swimming motility is particularly rapid at tens to hundreds of micrometres per s (ref. 22). This supports our hypothesis that microzones most probably occur in the thermocline, as phytoplankton sinking rates are reduced in this region²³. Also, Woods^{14,24} reports that sheets remain turbid while the surrounding water is clear, suggesting that particles are concentrated in the sheets.

The long eddy lifetimes in reduced turbulence may be important to other small-scale processes, such as formation of micro-scale nutrient patches²⁵. Longer eddy lifetimes may allow chemotactic swarming of motile phytoplankton around a zooplankton or nutrient particle. This potential behavioural response may explain how Lehman and Scavia's²⁵ results were possible despite Jackson's²⁶ calculations predicting otherwise.

In summary, generous estimates biased toward high nutrient concentrations show that: (1) significant organic enrichment extends less than 1 mm from a phytoplankter, and (2) the highest concentration, at the cell surface, is enhanced less than 10 times above background. Estimates from the more typical conditions yield apparently negligible microzones. Furthermore, bacteria cannot keep up with phytoplankton moving faster than the bacterial chemotactic swimming rate. The swimming speeds of motile phytoplankton, and average sinking rates of non-motile forms within the mixed layer, appear to preclude clustering. We suggest that clustering is limited to non-motile forms such as diatoms and seems most likely, and perhaps best studied, in the thermocline. There, hydrodynamic fine structure contributes to reduced phytoplankton sinking rates and provides an environment more amenable to maintenance of ordered, small-scale interactions between microbes than is present in the turbulent mixed layer. This may lead to a stratified microbial community across and within the thermocline.

We thank F. Azam, S. Brownlee, M. Scranton, C. Monaco and V. Gerard for comments on the manuscript; and J. Lumley, M. Legier, M. Silver, C. Decker, G. McManus, T. Bell and J. Hudak for discussion. The work was supported by a Jesse Smith-Noyes Fellowship (to J.G.M.), and by NSF grants OCE 8214498 and OCE 8207523.

Received 26 September 1984; accepted 18 April 1985.

1. Fuhrman, J. A. & Azam, F. *Mar. Biol.* **66**, 109–120 (1982).
2. Bell, W. H. & Mitchell, R. *Biol. Bull.* **143**(2), 265–277 (1972).
3. Azam, F. & Hodson, R. E. *Mar. Ecol. Prog. Ser.* **6**, 213–222 (1981).
4. Azam, F. & Ammerman, J. W. *Flows of Energy and Nutrients in Marine Ecosystems* (ed. Fasham, M. J. R.) (Plenum, New York, 1984).
5. Sharp, J. J. *Limnol. Oceanogr.* **22**(3), 381–399 (1977).
6. Lancelot, C. *Mar. Ecol. Prog. Ser.* **1**, 179–186 (1979).
7. Parsons, T. R. *et al. Biological Oceanographic Processes* (Pergamon, New York, 1977).
8. Chrost, R. H. & Faust, M. A. *J. Plankton Res.* **5**(4), 477–493 (1983).
9. Fogg, G. E. *Physiology and Biochemistry of Algae* (ed. Lewin, R. A.) (Academic, New York, 1963).
10. Carslaw, H. S. & Jaeger, J. C. *Conduction of Heat in Solids* (Oxford University Press, 1959).
11. Carlucci, A. F., Craven, D. B. & Henrichs, S. M. *Appl. envir. Microbiol.* **48**, 165–170 (1984).
12. Mopper, K. *et al. Mar. Chem.* **10**, 55–66 (1980).
13. Weast, R. C. (ed.) *Handbook of Chemistry and Physics* 56th edn (CRC Press, Cleveland, 1975).
14. Woods, J. D. *Met. Mag., Lond.* **97**, (1148), 65–72 (1968).
15. Berg, H. C. & Purcell, E. M. *Biophys. J.* **20**, 193–219 (1977).
16. DeLisi, C. *Nature* **289**, 322–323 (1981).
17. McNab, R. M. & Koshland, D. E. Jr. *Proc. natn. Acad. Sci. U.S.A.* **69**, 2509–2512 (1972).
18. Berg, H. C. *Nature* **254**, 389–392 (1975).
19. Adler, J. *Science* **166**, 1588–1597 (1969).
20. Dahlquist, F. W. *et al. Nature new Biol.* **236**, 120–123 (1972).
21. Bienfang, P. K. *Mar. Biol.* **61**, 69–77 (1980).
22. Lee, R. E. *Phycology* (Cambridge University Press, 1980).
23. Hobson, L. A. & Lorenzen, C. J. *Deep Sea Res.* **19**, 297–306 (1972).
24. Woods, J. D. & Fosberry, G. G. *Underwat. Ass. Rep.* **5**, 19 (1966–67).
25. Lehman, J. T. & Scavia, D. *Science* **216**, 729–730 (1982).
26. Jackson, G. A. *Nature* **284**, 439–441 (1980).
27. Batchelor, G. K. *The Theory of Homogeneous Turbulence* (Cambridge University Press, 1956).
28. Ozmidov, R. V. *Izv. Akad. Nauk. SSR* **1**, 439–448 (1965). (Engl. transl. *Izvestiya atmos. Oceanic Phys.* **1**, 257–261 (1965).
29. Gregg, M. C. *J. phys. Oceanogr.* **14**, 588–711 (1984).

Female mimicry in garter snakes

Robert T. Mason & David Crews

Institute of Reproductive Biology, Department of Zoology, University of Texas at Austin, Austin, Texas 78712, USA

In many diverse taxa, males of the same species often exhibit multiple mating strategies¹. One well-documented alternative male reproductive pattern is 'female mimicry', whereby males assume a female-like morphology^{2–8} or mimic female behaviour patterns^{9,10}. In some species males mimic both female morphology and behaviour^{2,7,11,12}. We report here female mimicry in a reptile, the red-sided garter snake (*Thamnophis sirtalis parietalis*). This form of mimicry is unique in that it is expressed as a physiological feminization. Courting male red-sided garter snakes detect a female-specific pheromone and normally avoid courting other males. However, a small proportion of males release a pheromone that attracts other males, as though they were females. In the field, mating aggregations of 5–17 males were observed formed around these individual attractive males, which we have termed 'she-males'. In competitive mating trails, she-males mated with females significantly more often than did normal males, demonstrating not only reproductive competence but also a possible selective advantage to males with this female-like pheromone.

On emergence from the winter den, female red-sided garter snakes are usually courted by 10–20, but sometimes as many as 100 males simultaneously, although only one male mates with the female¹³. Males begin to court females when rapid tongue-flicks by the male deliver pheromone cues from the female's dorsal surface to the male's vomeronasal organ^{14,15}. During the spring of 1983, a census was taken of presumed 'mating balls' at a study site in central Manitoba, Canada. In 33 of the observed 200 (14%) mating balls studied, no female was found; instead a single male was being courted by the other males. These

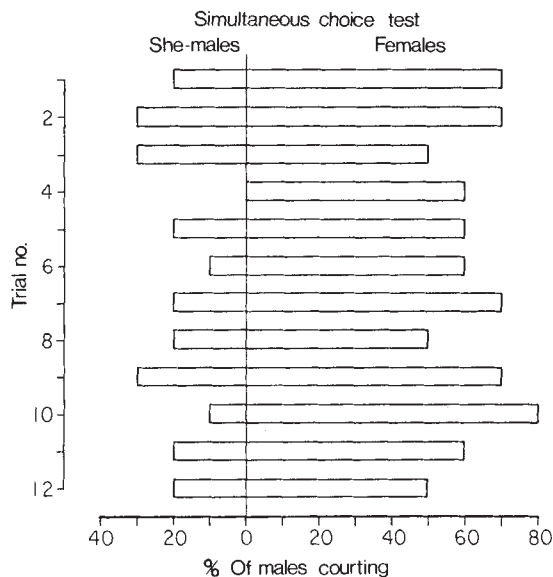


Fig. 1 Results of a simultaneous choice test. In this test a female and a she-male were placed in an arena with 10 courting males. After 5 min the number of males that courted either stimulus animal was recorded.

courted males (she-males) possessed normal hemipenes, testes and accessory sex structures and their mean snout-vent length and body weight did not differ significantly from those of other males taken at random from the population.

The attractiveness of she-males is not caused by prior physical contact with recently emerged females. Males rubbed against a recently emerged, attractive female were not courted by other males, agreeing with field and laboratory observations in which recently mated males that have been in contact with females for up to 1 h have never been observed to be courted. The attractiveness of the she-males appears to be as permanent as that of sexually mature females¹⁶. All of the she-males tested ($n=33$) elicited courtship from sexually active males when tested at 3, 18 and 40 weeks after capture.

Steroid hormone levels were assayed on blood samples taken in the field from males, females and she-males¹⁷. Analysis of variance revealed that females had significantly higher circulating levels of oestradiol than either males or she-males ($P<0.05$, Student-Newman Keuls) (Table 1). Although testosterone levels in males and females were not significantly different, she-males had significantly higher levels of testosterone than either females or males ($P<0.05$, Student-Newman Keuls). There were no significant differences in circulating levels of dihydrotestosterone. Presumably, the high levels of testosterone in the she-males was a result of either higher secretion or reduced clearance from the circulation.

Table 1 Circulating levels of sex steroid hormone in the two male morphs and the female red-sided garter snake

		Hormone (ng ml ⁻¹)		
	(n)	Testosterone	Oestradiol	Dihydro-testosterone
Male	9	5.720 ± 1.655	0.061 ± 0.012	0.231 ± 0.066
Female	9	0.233 ± 0.082	7.703 ± 3.440	0.093 ± 0.017
She-male	9	19.785 ± 5.703	0.075 ± 0.013	0.142 ± 0.025
<i>F</i> ratio (d.f. 2, 24)		7.68	3.69	2.45
<i>F</i> ≥ 3.40, <i>P</i> > 0.05				

Plasma levels of dihydrotestosterone, oestradiol and testosterone were determined by steroid radioimmunoassay. Values are the mean ± s.e.m.

Table 2 Transfer of the attractiveness pheromone

Trial no.	Female	Male	She-male
1	+	—	+
2	+	—	+
3	+	—	+
4	+	—	—
5	+	—	—

Ten sexually active male red-sided garter snakes were placed in an arena at the den site. Pheromone samples obtained from the skin surface of females, males and she-males were then presented to the test males. A sample was scored as attractive if five or more males courted for 10 s. Ten different males were used in each trial.

The pheromone produced by the she-males is apparently indistinguishable from that produced by females. Samples obtained from the skin surface using the non-polar solvent toluene revealed that the skin of females and she-males, but not of males, elicited courtship behaviour from conspecific males at the den site (Table 2). However, in simultaneous choice tests ($n=12$) in which 10 courting males were able to choose between a female and a she-male, the majority of males preferred the female in every case, although in each trial some males also courted the she-male (Fig. 1). Thus, females may produce more of the pheromone or they may produce a qualitatively different pheromone from that of she-males.

Competitive mating trials were conducted in the field in which groups of five males and five she-males were selected at random and released into an arena (1 m × 1 m) with an unmated, attractive female. In a total of 42 mating trials, 29 resulted in a she-male mating with the female and 13 with a male mating; thus, she-males mated significantly more often than did normal males ($\chi^2=11.85$, $P<0.001$). In field observations, she-males joining an established mating ball caused courting males to ignore the female and begin to court the she-male. A possible advantage to the she-male's attractiveness thus lies in an ability to gain a better position in the mating ball by confusing other males.

Currently, breeding experiments and future field work are being designed to resolve an enigma: why are all males not she-males if she-males do indeed have a reproductive advantage over normal males?

We thank Jim Blank, Jim Bull, Michael Moore, Eric Pianka, Mike Ryan and Joan Whittier for discussion; David Hinkley, Jim Bull and Fred Marsteller for statistical assistance; and Maire Anne Diamond and Yuki Morris for technical assistance. Bill Koonz and Merlin Shoesmith of the Manitoba Department of Natural Resources provided permits and contributed immensely to the collection of both animals and data during the field seasons. This work was supported by NIH grant NICHD 16687 and by a National Institute of Mental Health Research Scientist Development Award to D.C.

Received 14 February; accepted 16 April 1985.

- Halliday, T. *Sexual Strategy* (Oxford University Press, 1980).
- Gross, M. R. & Charnov, E. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6937-6940 (1980).
- Morris, D. *Behavior* **4**, 232-261 (1952).
- Gross, M. R. *Can. J. Zool.* **57**, 1507-1509 (1979).
- Gross, M. R. *Z. Tierpsychol.* **60**, 1-26 (1982).
- Dominey, W. *Nature* **284**, 546-548 (1980).
- Gross, M. R. in *Fish Reproduction: Strategies and Tactics* (eds Wootton, R. J. & Potts, G.) 55-75 (Academic, London, 1984).
- Gross, M. R. *Nature* **313**, 47-48 (1985).
- Thornhill, R. *Science* **205**, 412-414 (1979).
- Arnold, S. J. *Z. Tierpsychol.* **42**, 247-300 (1976).
- Dominey, W. *Am. Zool.* **24**, 385-396 (1984).
- Weldon, P. J. & Burghardt, G. M. *Z. Tierpsychol.* **65**, 89-102 (1984).
- Nobel, G. K. *Bull. Am. Mus. nat. Hist.* **73**, 673-725 (1937).
- Crews, D. *Horm. Behav.* **7**, 451-460 (1976).
- Garstka, W. & Crews, D. *Science* **214**, 681-683 (1981).
- Bona-Gallo, A. & Licht, P. *Herpetologica* **39**(2), 173-182 (1983).
- Camazine, B., Garstka, W., Tokarz, R. & Crews, D. *Horm. Behav.* **14**, 358-372 (1980).

Related spotted hyaenas forage together but do not cooperate in rearing young

M. G. L. Mills

National Parks Board, Kruger National Park, Private Bag X402, Skukuza 1350, South Africa

Spotted hyaenas (*Crocuta crocuta*) cooperate little in raising their young¹, in contrast to other social carnivores²⁻¹¹. It has been contended that the average relatedness between adult spotted hyaenas in a clan is so low that there is no kin selection pressure in favour of helping to rear young¹². In the southern Kalahari, however, I found that there is a high degree of relatedness between the members of clans and that kin selection may be implicated in ways other than cooperation in raising young.

Two clans were intensively studied in the Kalahari Gemsbok (South Africa) and adjacent Gemsbok (Botswana) National Parks from January 1979 to April 1984. The animals were individually marked or were recognizable. The degrees of relatedness (r) between the members of the clans were calculated making the following assumptions: (1) Females suckled only their own cubs¹. (2) The degree of relatedness between the original three key adult females of the Kousaunt clan (Fig. 1), were equivalent to the mean of the known degrees of relatedness between 10 female pairs from the Kaspersdraai clan, that is 0.38. (3) The resident adult male in a clan at the time when cubs

were conceived was their father. When there were two adult males present (Kousaunt clan, May 1981 to August 1983; Fig. 1), the behaviour of all but one of the females (Guy) towards the second male (Necklace) suggested that they would not have allowed him to mate with them. Accordingly, the male Hans was assumed to have fathered all cubs born during this period, except for Guy's cub, which was not included.

Spotted hyaenas of the southern Kalahari live in clans of 5–20 individuals inhabiting territories of 600–1,900 km² (M.G.L.M., manuscript in preparation). All males ($n = 13$) left their natal territories at 2½ to 3½ years old, but only 1 of 8 females did so. The average degree of relatedness between 231 pairs of hyaenas of the Kousaunt clan, which averaged 17 individuals during the study, was 0.29 and between 32 pairs of the Kaspersdraai clan, which averaged eight individuals, was 0.33.

The hyaenas' food consists primarily of large and medium-sized ungulates, 55.3% ($n = 197$) of which they killed. Observations of spotted hyaenas hunting gemsbok (*Oryx gazella*) calves (43% of kills), sub-adults and adults (10% of kills) and blue wildebeest (*Connochaetes taurinus*) of all ages (15% of kills), revealed no correlation between hunting success and hunting group size (Fig. 2). Single hyaenas, however, were unable to kill larger gemsbok, two and four being the smallest groups observed to catch sub-adults and adults, respectively. It was usually impossible to differentiate between hunting attempts on wildebeest calves and adults, but group size did not appear to be a large factor in determining hunting success, as Kruuk¹ also concluded. Furthermore, carrion, which is usually located by scent, can be located as efficiently by a single hyaena as by

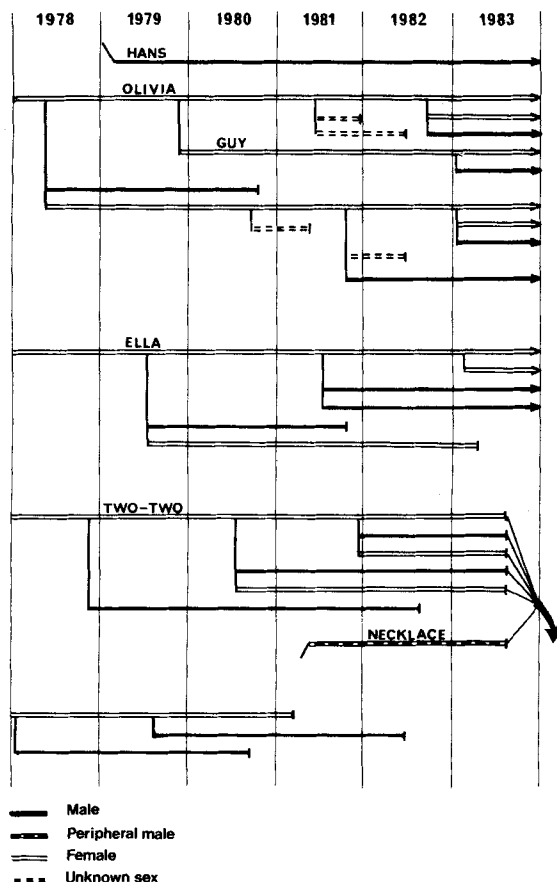


Fig. 1 Summary of the reproductive history of the Kousaunt clan from 1978 to 1983, showing the three coalitions and the subsequent fission of the clan. The birth of a litter is represented by a vertical line descending from the female's line; immigration of an individual is indicated by an oblique line. The ending of a horizontal line by one short vertical line denotes either death or emigration. The names of animals mentioned in the text are given.

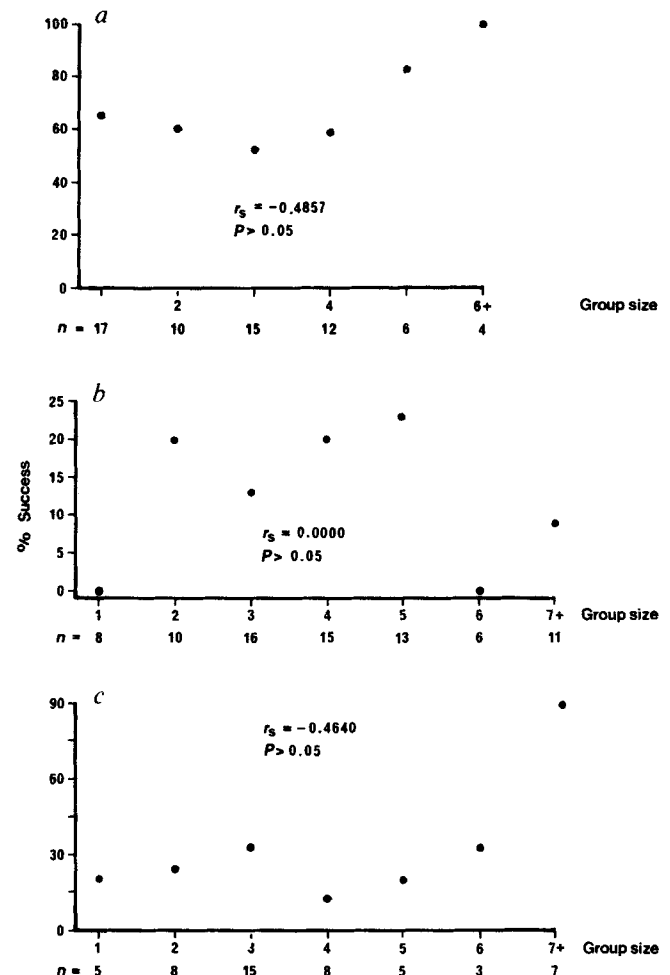


Fig. 2 The relationship between spotted hyaena hunting group size and hunting success of a, gemsbok calves; b, gemsbok sub-adults and adults; and c, wildebeest of all ages.

several. Of 566 foraging groups, 29.9% comprised one, 19.3% two, 18.7% three, 13.3% four, and 19.2% five or more animals. Foraging groups sizes, therefore, were often larger than needed for maximum foraging efficiency, so cooperation in hunting may not be the only selective pressure for group foraging in these spotted hyaenas.

Another reason for group foraging could be that larger groups are more successful at defending their food from competitors^{13,14}. In the southern Kalahari, however, spotted hyaenas and their chief competitors, lions (*Panthera leo*), are few. Only once did I observe other hyaenas (and on three occasions, lions) trying to appropriate a carcass from feeding spotted hyaenas ($n = 197$).

Could spotted hyaenas form these larger foraging groups to allow closely related individuals access to food, thus possibly enhancing their inclusive fitness? The degrees of relatedness between the members of 131 foraging groups of Kousaunt hyaenas and 28 foraging groups of Kaspersdraai animals, were compared with those for equal numbers of equal-sized foraging groups randomly generated by a computer from the pool of available hyaenas in each clan. In both cases the degrees of relatedness in the foraging groups were significantly higher than they were in the random groups (Kolmogorov-Smirnov one-sample test: Kousaunt, $D = 0.227$, $P < 0.01$, two-tailed test; Kaspersdraai, $D = 0.149$, $P < 0.01$, two-tailed test). These hyaenas, therefore, showed a strong tendency to preferentially forage with their close kin.

Certain individuals were often excluded from food. I compared the degrees of relatedness of only those members of 32 feeding groups from the Kousaunt clan that were permitted to eat at a given carcass, with those of 32 random groups from the clan, chosen as described above. Here again the degrees of relatedness were significantly higher in the observed groups than in the random groups (Kolmogorov-Smirnov one-sample test, $D = 0.088$; $P < 0.01$, two-tailed test).

In 1983 a split occurred in the expanding Kousaunt clan (Fig. 1). Three coalitions comprising the three surviving original adult females plus their offspring (Fig. 1) developed in the clan (unpublished observations); this reached a climax in August 1983 when all the members of the Olivia and Two-Two coalitions were observed to engage in a fierce intra-clan fight. The battle probably was not their first as, prior to this, two members of the Olivia coalition had sustained injuries. After the battle the Two-Two coalition, together with the male (Necklace) (Fig. 1) departed and settled in an adjoining area. The Olivia and Ella coalitions continued to inhabit the same area and den communally.

Relatedness between individuals, therefore, is more important in spotted hyaena society than was thought previously. Apart from sharing dens, why then do they not cooperate in raising young as brown hyaenas (*Hyaena brunnea*) do^{10,11}? Regurgitation of food for young is not seen in the Hyaenidae, probably because of the rapid rate at which food passes from the stomach to the intestines¹. Food carrying, which is observed in the brown hyaena, would be difficult to accomplish in spotted hyaenas¹⁵ as: (1) much of their food consists of large carcasses, which the hyaenas often have to travel long distances to obtain; (2) the social nature of their feeding habits¹ means that there is often nothing suitable left to take back to a den; and (3) at the communal dens the larger cubs would get most of the food. Cubs, therefore, rely heavily on their mother's milk during their first year¹; this imparts a heavy nutritional load on the females, which may only be able to afford to suckle their own offspring. Finally, the small den tunnels^{1,16,17} provide ideal protection for cubs when the adults are away, thus precluding the need to leave guards. I suggest that spotted hyaenas cooperate little in raising their young because of ecological pressures and not because of low levels of relatedness between clan members.

I thank the National Parks Board for logistic support, Michael Knight and Hermanus Jaegers for help in the field, Peter Retief for help with statistics and Drs David Macdonald and Anthony Hall-Martin for comments.

Received 19 February; accepted 25 April 1985.

1. Kruuk, H. *The Spotted Hyena, a Study of Predation and Social Behaviour* (University of Chicago Press 1972).
2. Kühme, W. Z. *Tierpsychol.* **22**, 495-541 (1965).
3. Mech, L. D. *The Wolf* (Natural History Press, New York, 1970).
4. Schaller, G. B. *The Serengeti Lion, a Study of Predator-Prey Relations* (University of Chicago Press, 1972).
5. Bertram, B. C. R. *J. Zool.* **177**, 462-482 (1975).
6. Rood, J. P. *Nature* **248**, 176 (1974); *Z. Tierpsychol.* **48**, 277-287 (1978); *Anim. Behav.* **28**, 143-150 (1980).
7. Camenzind, F. J. in *Coyotes: Biology, Behaviour and Management* (ed. Bedkoff, M.) 267-294 (Academic, New York, 1978).
8. Macdonald, D. W. *Nature* **282**, 69-71 (1979).
9. Moehlman, P. D. *Nature* **277**, 382-383 (1978).
10. Mills, M. G. L. *Behav. Ecol. Sociobiol.* **10**, 131-136 (1982).
11. Owens, D. D. & Owens, M. J. *Nature* **308**, 843-845 (1984).
12. Bertram, B. C. R. in *Serengeti, Dynamics of an Ecosystem* (eds Sinclair, A. R. E. & Norton-Griffiths, M.) 221-248 (University of Chicago Press, 1979).
13. Caraco, R. & Wolf, L. L. *Am. Nat.* **109**, 343-352 (1975).
14. Lamprecht, J. Z. *Tierpsychol.* **46**, 337-343 (1978).
15. Mills, M. G. L. *Z. Tierpsychol.* **63**, 331-342 (1983).
16. Watson, R. M. E. *Afr. Wildlife J.* **122**-123 (1965).
17. Henschell, J. R., Tilson, R. & von Blotnitz, S. *Afr. Archaeol. Bull.* **34**, 127-131 (1979).

Defective ability to self-renew *in vitro* of highly purified primitive haematopoietic cells

Elaine Spooncer, B. I. Lord & T. M. Dexter

Paterson Laboratories, Christie Hospital and Holt Radium Institute, Withington, Manchester M20 9BX, UK

Stromal cells play a critical role in haematopoiesis, both in a permissive and, probably, in a directive manner¹⁻⁵. Study of the interactions between stromal cells and haematopoietic stem cells, however, is difficult to perform using whole bone marrow, in which stem cells are indistinguishable from precursor cells and maturing haematopoietic cells, and where stromal and haematopoietic cells co-exist in a heterogeneous mixture. We have purified primitive haematopoietic spleen colony-forming cells (CFU-S) using fluorescence-activated cell sorting (FACS) and produced CFU-S populations which approach 100% purity (ref. 6 and B.I.L. and E.S., in preparation). This cell population is devoid of significant stromal cells and mature haematopoietic cells. Here, we report that when purified CFU-S are seeded onto a stromal adherent layer *in vitro*, foci of haematopoietic cells develop within the stroma followed by production of a wave of maturing and mature progeny. However, self-renewal of CFU-S does not occur and haematopoietic activity rapidly declines, indicating that caution should be applied in the use of highly purified stem cells for human bone marrow transplantation.

Irradiated long-term marrow culture adherent layers were used as the source of inductive haematopoietic stroma². We seeded the stroma (or empty culture flasks) with various numbers of normal marrow cells (Table 1). Normal bone marrow cells (5×10^5 containing 130 CFU-S) plated into an empty tissue-culture flask produced two small stromal cell foci per flask and no haematopoietic cells over 7 weeks. The same number of normal bone marrow cells plated onto an irradiated adherent layer led to the production and release of CFU-S into the growth medium for several weeks. Note that this is an underestimate of the total CFU-S as this does not include the CFU-S lodged in the adherent cell layer⁷. It is not known why haematopoiesis was exhausted in these cultures after 7 weeks. A larger inoculum of normal bone marrow cells, 5×10^6 cells, used to initiate a long-term bone marrow culture adherent layer, produced a confluent stromal adherent layer that promoted the maintenance of relatively few CFU-S for a limited duration, as described previously⁸. However, when 5×10^6 normal bone marrow cells were seeded onto an irradiated adherent layer, many CFU-S were produced for several weeks and the duration of haematopoiesis was clearly extended. In agreement with previous studies, fully competent long-term bone marrow cultures could be established from 1.5×10^7 normal bone marrow cells

Table 1 CFU-S production in irradiated marrow stromal cell layers and control flasks inoculated with normal marrow cells

Input normal bone marrow	Input CFU-S	Irradiated stromal cells	CFU-S per culture (release to growth medium) (weeks)						
			1	2	3	4	5	6	7
5×10^5	130	No	0	0	0	0	0	0	0
5×10^6	1,300	No	91	20	30	20	10	0	13
1.5×10^7	13,000	No	700	386	432	576	580	624	394
0	0	Yes	0	0	0	0	0	0	0
5×10^5	130	Yes	18	39	101	40	92	16	0
5×10^6	1,300	Yes	139	250	650	250	130	ND	100

To establish irradiated marrow stromal cell layers, we set up long-term bone marrow cultures as described previously¹. Briefly, each culture was initiated by flushing the contents of one femur (1.5×10^7 cells) from (C57BL/6 $\times$ DBA/2) F₁ donors into 10 ml Fischer's medium (Gibco) supplemented with horse serum (20% v/v) (Flow Labs) and 10^{-6} M hydrocortisone sodium succinate. Cultures were fed weekly by removing 5 ml growth medium and adding 5 ml fresh medium. After 4 weeks the cultures were irradiated with 1,500 rad ¹³⁷Cs, followed by a total change of medium 6 h later. The dose kills all haematopoietic cells yet spares relatively quiescent stromal cells and provides a nude inductive stroma that can function to promote the growth of haematopoietic cells for at least several weeks². The stromal cell layers were gently rinsed with Fischer's medium 7–10 days after irradiation and the appropriate number of fresh bone marrow cells was added in 10 ml Fischer's medium with 20% v/v horse serum. At weekly intervals, cultures were fed and the suspension cells collected during feeding were diluted (if necessary) and assayed for CFU-S according to the method of Till and McCulloch¹². Recipient mice were killed after 10 days. In some experiments, fresh bone marrow cells were seeded into empty culture flasks. ND, not done.

Table 2 CFU-S production in irradiated marrow stromal cell layers inoculated with FACS-sorted bone marrow cells

Input FACS-BM	Input CFU-S	Irradiated stromal cells	CFU-S per total culture (weeks)		
			1	2	3
10,000	700	Yes	—	19	—
5,000	350	Yes	—	—	0
2,500	175	Yes	22.5	8.5	—
1,250	87	Yes	—	—	0
625	44	Yes	9	2	4
312	22	Yes	—	—	4
20	1.3	Yes	—	—	0
500	26	No	0	0	—

Irradiated long-term bone marrow cultures were established as described in Table 1 legend. Normal bone marrow cells were enriched for CFU-S content by a modification (B.I.L. and E.S., in preparation) of the electronic sorting technique described by Visser *et al.*⁶. Normal bone marrow cells were labelled with $2 \mu\text{g ml}^{-1}$ fluorescein isothiocyanate-conjugated wheat-germ agglutinin (WGA-FITC) and subjected to a density cut in metrizamide. Cells with density $\leq 1.080 \text{ g ml}^{-1}$ were collected, washed in phosphate-buffered saline and 'sorted' on a fluorescence-activated cell sorter (FACS IV; Beckton & Dickinson). Electronic windows were set to select highly WGA-FITC-labelled cells with moderate forward and low perpendicular light scatter. The selected cells were collected in Fischer's medium containing 10% horse serum, washed and re-sorted under the same criteria as before. WGA-FITC was removed from the re-sorted cells by incubating for 15 min at 37 °C in isotonic, 0.2 M *N*-acetyl D-glucosamine⁶ before assaying the processed cells for CFU-S content¹². Recipient mice in the CFU-S assay were killed after 12 days. The FACS-BM contained 70–90 CFU-S per 10^3 cells. In accordance with the measured seeding efficiency of the FACS-BM cells to the spleen (data not shown), the concentration of CFC-S was calculated to be 50–107%. The appropriate number of FCS-BM cells were seeded onto the irradiated stromal cell layers, cultures terminated at various times and the adherent cells scraped off the flask and pooled with the suspension cells. One-fifth of each culture was injected into five irradiated mice to measure total CFU-S per culture. The results of one experiment are shown here, but we have obtained identical results in four other experiments.

with subsequent production of high numbers of CFU-S for many weeks¹. From these data we can conclude that an inoculum of normal bone marrow containing ~ 130 CFU-S is capable of colonizing a haematopoietic environment and maintaining self-renewal for at least 4–5 weeks. There does not seem to be appreciable stromal cell contribution to the irradiated stromal elements by the fresh normal bone marrow cells at this dose.

When various numbers of FACS-treated bone-marrow cells (FACS-BM cells) were seeded onto irradiated adherent layers and the cultures were terminated after 1,2 or 3 weeks to measure

Table 3 Generation of putative haematopoietic cell foci in irradiated marrow stromal cell layers

Input FACS-BM	Foci of putative haematopoietic cells (weeks)			
	1	2	3	4
5,000	+	++	—	—
1,250	+	++	—	—
312	+	+	+	—
78	—	+/-	+	+/-
20	—	+/-	+	+/-
0	—	—	—	—

Foci present in 60–100% random fields of view (++); 30–60% (+); 1–30% (+/-); no field of view (—). Irradiated marrow stromal cell layers were seeded with FACS-BM exactly as described in Table 2 legend. Cultures were examined weekly using a Leitz diavert inverted microscope. Foci of putative haematopoietic cells were classified according to the scheme above.

Table 4 CFU-S production in irradiated marrow stromal cell layers inoculated with reconstituted FACS-BM cells

Input	Input CFU-S	CFU-S per total culture (weeks)			
		1	2	3	4
NBM 5×10^5	130	34	240	174	132
Reconstituted FACS-BM 6.5×10^5	180	>250	>250	ND	185

Irradiated marrow stromal cell layers were established, seeded with cells and assayed weekly for total CFU-S per culture as described in Tables 1 and 2 legends. In one set of cultures, 5×10^5 normal bone marrow cells were seeded onto irradiated marrow stromal cell layers. For another set, the bone marrow was sorted through the FACS as described in Table 2 legend. After sorting, the FACS-BM was recombined with cells rejected at the metrizamide stage and during the FACS sorting to produce reconstituted FACS-BM cells, which were seeded onto irradiated marrow stromal cell layers.

the total CFU-S present in the growth medium and within the adherent layer, we consistently detected few or no CFU-S (Table 2). Apparently the FACS-BM CFU-S are neither maintained nor produced in such cultures, even when an inoculum of 700 CFU-S was used, equivalent to $\sim 3 \times 10^6$ normal bone marrow cells. Although these cultures were devoid of CFU-S, survival and production of maturing haematopoietic cells occurred. When the FACS-BM cells were plated onto the stromal cells, clonal foci of haematopoietic cells developed in the adherent layer. The number of foci was roughly proportional to the

number of FACS-BM cells introduced to the cultures and their sequential development from two or four cell foci into typical 'cobblestone' areas composed of hundreds of cells was observed. Furthermore, these foci were not stable and we observed the appearance, growth and disappearance of individual foci over a period of days (Table 3). Later than this, no foci were seen and the adherent layer became infiltrated randomly by macrophage-type cells. The cells in these foci are clearly not CFU-S; granulocyte-macrophage colony-forming cell assays (data not shown) only detected many macrophage cluster-forming cells. However, at the time of peak cobblestone formation the cell counts in the growth medium were often $>10^5$ per ml and the cells being produced consisted of blast cells, immature and maturing neutrophils and macrophages. Because of this, it is likely that the foci consist of maturing myeloid cells, although the identity of the blast cells is unknown.

From these data it is clear that normal bone marrow cells (containing limited numbers of CFU-S) can maintain haematopoiesis for some weeks, whereas FACS-BM (containing the same number of, or more, CFU-S) shows no ability to self-renew CFU-S or to establish long-term haematopoiesis *in vitro*. However, the FACS-BM cells can 'colonize' the stroma and form foci of cells of haematopoietic origin that undergo a limited amount of proliferation and development. The inability of FACS-BM cells to express self-renewal capacity does not result from the physical and mechanical stress of FACS sorting and the binding and removal of wheat-germ agglutinin as shown in Table 4.

Thus, other possibilities must be considered for the lack of growth of FACS-sorted CFU-S in irradiated long-term bone marrow cultures: (1) Irradiated stroma is defective in its ability to promote self-renewal of CFU-S. This seems unlikely because haematopoiesis can be established on the irradiated stroma by 5×10^5 normal bone marrow cells (which contain few stromal cells). (2) Removal of some accessory cell(s) important for the maintenance or control of CFU-S self-renewal. We are currently investigating this possibility. (3) Lack of feedback inhibition on differentiation. As FACS-BM cells differ from normal marrow cells in the absence of mature cells, it is possible that the mature cells somehow regulate the probability of self-renewal of CFU-S. (4) Finally, it is possible that the CFU-S are not the cells which repopulate haematopoiesis in long-term bone marrow cultures, that is, they are not the most primitive stem cells. Although FACS-purified CFU-S can promote 30-day survival of irradiated mice, no study has been made of the longer-term quality of the CFU-S⁶. The haematopoietic stem-cell compartment is clearly a heterogeneous population^{2,9} and certain elements of the CFU-S population (or even primitive stem cells that do not have spleen colony-forming capacity) may be discarded by the sorting procedure. Thus, although FACS-sorting seems to select for CFU-S with an average (or high) 'self-renewal' capacity *in vivo* in the short term⁶, the CFU-S may still not be representative of the stem cells.

These findings have obvious clinical importance. Recently, various methods for cleaning up human bone marrow cells before transplantation have been described, aimed at removing from the graft T cells^{10,11} responsible for provoking graft-versus-host disease. Treatment of bone marrow grafts in this way usually leads to alleviation of the disease, but also leads to a surprisingly high level of graft failure in some patients, for unknown reasons¹¹. Therefore, an understanding of why FACS-BM CFU-S do not self-renew *in vitro* may explain some of the problems associated with engraftment of human marrow *in vivo*. Our results indicate caution in the use of highly purified human haematopoietic progenitor cells for transplantation. Furthermore, the observation that CFU-S from FACS-BM do not self-renew *in vitro* provides an opportunity to investigate circumstances and conditions required for the self-renewal of haematopoietic stem cells.

This work was supported by the Cancer Research Campaign, of which T.M.D. is a fellow. We thank Ruth Roberts for preparing some of the data.

Received 28 December 1984; accepted 3 May 1985.

1. Dexter, T. M., Spooncer, E., Simmons, P. & Allen, T. D. in *Long-Term Bone Marrow Culture* (eds Wright, D. G. & Greenberger, J. S.) 57-96 (Liss, New York, 1984).
2. Schofield, R. & Dexter, T. M. *Leukemia Res.* 9, 305-313 (1985).
3. Dexter, T. M. *J. cell. Physiol. Suppl.* 1, 87-94 (1982).
4. Wolf, N. S. *Clin. Haemat.* 8, 469-500 (1979).
5. Dexter, T. M. & Moore, M. A. S. *Nature* 269, 412-414 (1977).
6. Visser, J. W. M., Bauman, J. G. J., Mulder, A. H., Eliason, J. F. & de Leeuw, A. M. *J. exp. Med.* 159, 1576-1590 (1984).
7. Dexter, T. M., Spooncer, E., Varga, J., Allen, T. D. & Lanotte, M. in *Haematopoietic Stem Cells* (eds Killman, Sv-Aa, Cronkite, E. P. & Muller-Berat, C. N.) 303-322 (Munksgaard, Copenhagen, 1983).
8. Spooncer, E. & Dexter, T. M. *Biblioth. Haemat.* 48, 366-383 (1984).
9. Schofield, R. *Blood Cells* 4, 7-25 (1978).
10. Trigg, M. E. *et al. Expl. Hemat.* 12, 412 (1984).
11. Hows, J., Parreira, A., Yin, J., Goldman, J. & Gordon-Smith, E. C. *Exp. Hemat.* 12, 412 (1984).
12. Till, J. E. & McCulloch, E. A. *Radiat. Res.* 14, 213-222 (1961).

Expression of the *c-fms* proto-oncogene during human monocytic differentiation

Eric Sariban, Thomas Mitchell & Donald Kufe

Laboratory of Clinical Pharmacology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA

The McDonough strain of the feline sarcoma virus contains a transforming gene (*v-fms*) which contains partial nucleotide homology with proto-oncogenes encoding tyrosine kinases¹⁻⁴. One of the *v-fms*-encoded products, gp140^{fms}, is a cell surface transmembrane glycoprotein that may function as a growth factor receptor^{5,6}. Although *c-fms* transcripts have been detected in placental trophoblasts⁷⁻¹⁰ and normal human bone marrow¹¹, the role of the *c-fms* gene product is unknown. We now report that induction of monocytic, but not granulocytic, differentiation of human HL-60 leukaemic cells is associated with expression of *c-fms*, preceded by that of *c-myc* and *c-fos*. Because *c-fms* transcripts are also detectable in peripheral blood monocytes and in blasts from certain patients with myelomonocytic leukaemia, the *c-fms* gene product may play a role in monocytic differentiation.

The *c-fms* gene and the normal cellular homologue of the transforming gene of the FBJ murine osteosarcoma virus (*v-fos*) are transcriptionally active in the cells of mouse extra-embryonal membranes⁷⁻¹⁰. Expression of the *c-fms* gene was undetectable in a variety of postnatal mouse tissues⁷⁻⁹, whereas *c-fos* transcription was restricted to haematopoietic cells¹⁰. This specific pattern of transcriptional activity suggests that both *c-fos* and *c-fms* are involved in certain differentiation pathways^{7,9}. *c-fms* gene expression is associated with monocytic differentiation^{10,12,13} and *c-fms* transcripts have been detected in undifferentiated mouse haematopoietic cells¹², certain human tumours¹⁴ and normal human bone marrow¹¹. Moreover, the detection of *c-fms* sequences on chromosome 5 (ref. 15) and an association of the 5q syndrome with certain myelodysplastic syndromes¹⁶ and acute myelogenous leukaemia¹⁷ suggested that *c-fms* gene expression is related to haematopoietic cell differentiation.

We have used the human HL-60 promyelocytic cell line¹⁸ to study the role of *c-fms* expression during induction of haematopoietic cell differentiation. HL-60 cells mature morphologically and functionally towards granulocytes after induction with dimethyl sulfoxide (DMSO) or hexamethylene bisacetamide (HMBA)^{19,20}. In contrast, these cells can be induced to differentiate along a monocyte-macrophage pathway when treated with 12-O-tetradecanoylphorbol-13-acetate (TPA)²¹ and 1,25 dihydroxy-vitamin D₃ (1,25(OH)₂D₃)²². HL-60 cells were collected 24 and 72 h after addition of each of these inducers to monitor for *c-fms* expression. Although no *c-fms* RNA was detected in uninduced HL-60 cells (Fig. 1, lane 1), *v-fms* DNA hybridized with a 5-kilobase (kb) transcript detectable after 24 h of TPA and 1,25(OH)₂D₃ treatment. In contrast, *c-fms* RNA was not detected following induction of HL-60 cells with DMSO and HMBA. Similar results were obtained after

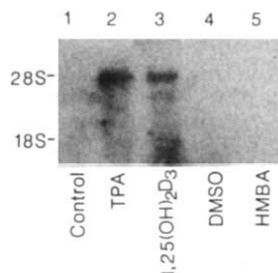


Fig. 1 Level of *c-fms* RNA during induction of HL-60 monocytic (TPA and $1,25(\text{OH})_2\text{D}_3$) and granulocytic (DMSO and HMBA) differentiation. HL-60 cells were grown as described previously¹⁸ in the presence of 3.3×10^{-8} M TPA (Sigma), 5×10^{-7} M $1,25(\text{OH})_2\text{D}_3$ (Hoffman-LaRoche), 1.25% DMSO or 4 mM HMBA for 24 h. Total cellular RNA was purified by the guanidine thiocyanate-caesium chloride method²⁹, analysed by electrophoresis of 15 μg RNA through 1% agarose-formaldehyde gels followed by Northern blot transfer to nitrocellulose³⁰. The 1.0-kb *Pst*I fragment of the *v-fms* gene was purified from the pSM3 plasmid⁴ by preparative gel electrophoresis and nick-translated to specific radioactivities of $\sim 5 \times 10^8$ c.p.m. μg^{-1} DNA. The filters were pre-hybridized at 42 °C for 16–24 h in buffer consisting of 50% formamide, $5 \times \text{SSC}$, 0.1% SDS, $5 \times \text{Denhardt's}$ and 200 $\mu\text{g ml}^{-1}$ salmon-sperm DNA. The RNA blots were then hybridized for 16–24 h at 42 °C with 2×10^6 c.p.m. nick-translated probe per ml hybridization buffer (same as the pre-hybridization buffer, with $1 \times \text{Denhardt's}$). After hybridization, the blots were washed twice in $2 \times \text{SSC}$, 0.1% SDS at room temperature for 60 min and then washed twice in $0.1 \times \text{SSC}$, 0.1% SDS at 55 °C for 30 min. The blots were then dried and exposed to X-ray film with an intensifying screen at -70 °C. Cytocentrifuge smears of HL-60 cells were examined after 3 days of continuous drug exposure for α -naphthyl acetate esterase (NSE) activity and nitroblue tetrazolium (NBT) reduction³¹. Adherence to the tissue culture flask was also monitored on day 3. Control HL-60 cells: 0% adherence, $13.3 \pm 1.5\%$ NSE positive, $7.3 \pm 1.5\%$ NBT positive; TPA-treated cells: $84.7 \pm 1.8\%$ adherence, $71.3 \pm 2.9\%$ NSE, $13.3 \pm 2.4\%$ NBT; $1,25(\text{OH})_2\text{D}_3$ -treated cells: 0% adherence, $86.0 \pm 5.6\%$ NSE, $99.3 \pm 0.8\%$ NBT; DMSO-treated cells: 0% adherence, $19.3 \pm 3.3\%$ NSE, $94.3 \pm 1.9\%$ NBT; HMBA-treated cells: 0% adherence, 0% NSE, $84.5 \pm 3.4\%$ NBT. (Results are expressed as mean $\pm$ s.d. of four determinations).

72 h treatment with each inducer and *c-fms* transcripts were not detected (data not shown) after TPA exposure of an HL-60 blast subclone²³ that does not undergo differentiation following treatment with inducers of monocyte or granulocyte differentiation. Taken together, these findings suggest that *c-fms* is induced when the HL-60 cells mature along the monocyte-macrophage pathway and that TPA is not a general inducer of *c-fms* gene transcription.

The temporal relationship between the addition of TPA and *c-fms* expression is shown in Fig. 2a. The *c-fms* transcripts appear 6 h after TPA addition, whereas maximum levels of *c-fms* RNA were present following 24–48 h of induction. These findings contrast with *c-myc* expression which, as shown previously²⁴, was not detected 24 h after TPA induction (Fig. 2a). Similar declines in *c-myb* transcripts occur following TPA treatment of HL-60 cells¹⁹, but these changes in *c-fms* and *c-myc* RNA during induction of HL-60 monocytic differentiation were not associated with altered *N-ras* or actin gene expression (Fig. 2a). We have also monitored *c-fms* expression during 5 days of TPA or $1,25(\text{OH})_2\text{D}_3$ treatment and find that *c-fms* RNA remains detectable following induction with $1,25(\text{OH})_2\text{D}_3$ (Fig. 2b). Similar results have been obtained after induction with TPA. Moreover, TPA treatment of HL-60 cells for 20 min is sufficient for commitment along the monocyte differentiation pathway¹³. The *c-fms* transcripts, however, were not detectable 24 h after a similar 20-min TPA exposure (data not shown).

Within 1 h of treating HL-60 cells with TPA, *c-fos* messenger RNA is detectable¹³. We have also examined the relationship

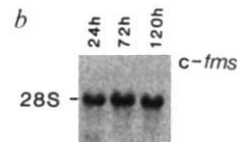
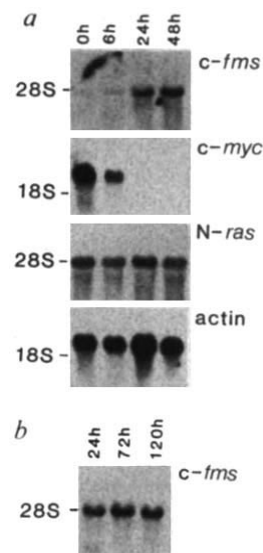


Fig. 2 Temporal relationship of *c-fms* expression during induction of HL-60 cells by TPA and $1,25(\text{OH})_2\text{D}_3$. **a**, HL-60 cells were treated with 3.3×10^{-8} M TPA for 6, 24 and 48 h. RNA (15 μg) was then hybridized to the ^{32}P -labelled nick-translated cDNA probes. The 1.6-kb *Clal*/*Eco*R1 fragment of the human *c-myc* 3' exon was purified from the pMC41-3 RC plasmid³². The 1.5-kb *p52C*-*Eco*R1 fragment of the HL-60 *N-ras* gene was purified from a reconstructed pBR322 plasmid³³. The 2.0-kb *Pst*I fragment of the chicken β -actin gene was purified from the pA1 plasmid³⁴. All methods are described in Fig. 1 legend. **b**, HL-60 cells were treated with 5×10^{-7} M $1,25(\text{OH})_2\text{D}_3$ for 24, 72 and 120 h. RNA (20 μg) was then hybridized to the ^{32}P -labelled *v-fms* probe.

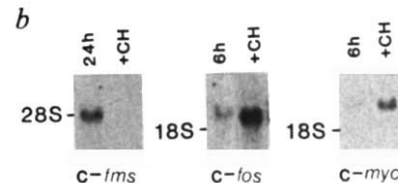
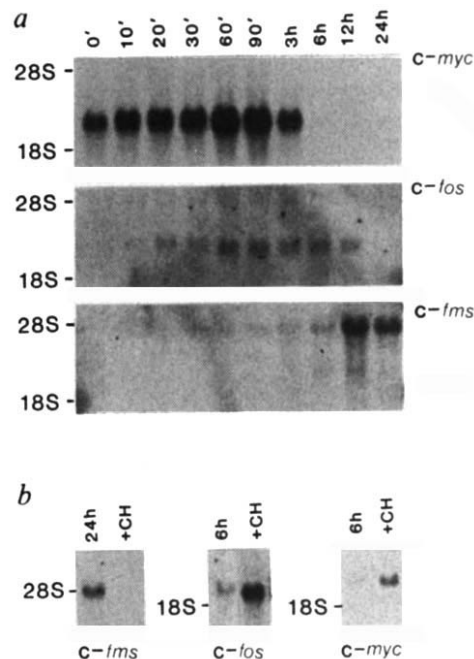


Fig. 3 **a**, Relationship between expression of *c-fms*, *c-fos* and *c-myc* in TPA-induced HL-60 cells; **b**, effect of protein synthesis inhibition on *c-fms*, *c-fos* and *c-myc* expression. **a**, HL-60 cells were treated with 3.3×10^{-8} M TPA and collected at the times indicated. Total cellular RNA (20 μg) and specific transcripts were detected by hybridization to ^{32}P -labelled complementary DNA probes as described in Figs 1, 2 legends. The 1-kb *Pst*I fragment of the *v-fos* gene was purified from the pfos-1 plasmid³⁵. **b**, HL-60 cells were treated for 6 or 24 h with 3.3×10^{-8} M TPA alone and in combination with cycloheximide (+CH). Procedures for RNA preparation and analysis were as described in Figs 1, 2 legends except that 20 μg total RNA were used in each lane.

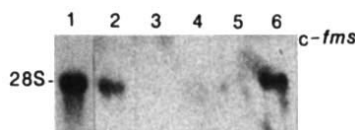


Fig. 4 *c-fms* gene transcripts in human monocytes and blasts from patients with myelomonocytic leukaemia. Peripheral blood monocytes and blasts from patients with myelomonocytic leukaemia were prepared as described previously²⁹. The ³²P-labelled *v-fms* probe was then hybridized to RNA (20 µg) obtained from these cell preparations: lane 1, normal peripheral blood monocytes (100% adherent cells); lane 2, TPA-induced HL-60 cells (24 h); lanes 3–6, blast preparations from four patients with myelomonocytic leukaemia (leukaemic blast composition by morphology: lane 3, 80%; lane 4, 79%; lane 5, 73%; and lane 6, 85%). All methods as described in Fig. 1 legend.

between changes in *c-myc*, *c-fos* and *c-fms* RNA during 24 h of TPA induction. The *c-myc* RNA level increased fourfold by 90 min after TPA addition (determined by optical densitometry of the autoradiograph) and then declined to nearly undetectable levels by 6 h (Fig. 3a). As reported previously¹³, *c-fos* transcripts accumulated within 20–30 min following induction, reached a maximum level at 90 min and declined thereafter. In contrast, *c-fms* expression occurred later, 6 h after TPA induction, and increased by 24 h (Fig. 3a). Clearly, the decrease in *c-myc* RNA and the increase of *c-fos* RNA both precede induction of *c-fms* expression.

We have extended our comparison of *c-myc*, *c-fos* and *c-fms* expression by monitoring the effects of cycloheximide on the level of these transcripts. Although *c-fms* expression was detectable in the 6-h TPA-treated cells, concurrent cycloheximide treatment inhibited the induction of *c-fms* RNA. Similar results were obtained after 24 h of TPA and cycloheximide treatment (Fig. 3b). In contrast, cycloheximide increased expression of both *c-myc* and *c-fos* (Fig. 3b). Thus, cycloheximide potentiates expression of the *c-myc* and *c-fos* genes known to be induced by mitogens^{13,25}, whereas our results suggest that induction of *c-fms* RNA requires new protein synthesis.

The differential sensitivity of *c-fos* and *c-fms* expression to cycloheximide, as well as the possible sequencing of these genes, may be related to specific functions of the gene products during monocyte proliferation and differentiation. Transcription of the *c-fos* gene, but not the *c-fms* gene, occurs within minutes of stimulating quiescent mouse fibroblast 3T3 cells with TPA²⁶. The early increase in both *c-myc* and *c-fos* mRNA, but not *c-fms* mRNA, has also been reported in TPA-treated quiescent fibroblasts and suggests that the *c-myc* and *c-fos* gene products are necessary for the mitogenic response²⁶. Conversely, transfection of F9 teratocarcinoma cells with an exogenous *c-fos* gene induces expression of only certain genes in the differentiation programme, thus indicating that maturation to a terminal stage of differentiation requires the expression of additional genes not regulated by *c-fos*²⁷. Thus, in TPA-treated HL-60 cells, the *c-fos*-encoded nuclear protein may play a dual role in both cell growth and initiation of differentiation. In contrast, the *c-fms* gene product may function in terminal differentiation along the monocyte-macrophage pathway.

Therefore, we have monitored the level of *c-fms* transcripts in circulating human monocytes and in blasts from four patients with myelomonocytic leukaemia (Fig. 4), and found that peripheral blood monocytes (Fig. 4, lane 1) and one of four leukaemic blast preparations had detectable levels of *c-fms* RNA (Fig. 4, lanes 3–6). The size of the transcript detected in leukaemic cells was identical to that observed in TPA-induced HL-60 cells (Fig. 4, lane 2) and in circulating monocytes. Furthermore, the 5-kb transcript was also detectable after TPA-induction of the human U-937 cell line to macrophages (data not shown). In contrast, *c-fms* RNA was undetectable in human granulocytes²⁸, peripheral blood T-cells, K562 erythroleukaemia

cells, B-cell lines (Daudi and JY), a T-cell line (MOLT-4) and AG1522 diploid fibroblasts.

Murine myeloid monocytic leukaemia WEHI-3B cells contain two *c-fms* transcripts of 4.1 and 8.4 kb (ref. 9). A smaller 3.7-kb *c-fms* transcript has been detected in mouse placental tissue⁸ and a transcript of the same size has been detected in a trophoblastic cell line (BeWo) derived from a gestational choriocarcinoma of the human fetal placenta⁹. In contrast, the results of the present study suggest that the *c-fms* transcript is 5 kb long in HL-60 and U-937 cells induced along a monocyte lineage, monocytes and certain leukaemic blasts. These differences in transcript size may be related to the physiological function of the *c-fms* encoded products in gestational and haematopoietic tissues.

Thus, expression of the *c-fms* gene has been observed in haematopoietic cell lines induced along the monocyte differentiation pathway. Furthermore, the detection of the *c-fms* transcripts in circulating monocytes, but not granulocytes or lymphocytes, implies that the gene product is involved specifically in the process of monocyte-macrophage differentiation. The localization of *c-fms* to chromosome 5 (ref. 15) is of interest in view of a specific deletion involving this region in refractory anaemia¹⁶ and acute myelogenous leukaemia¹⁷. Unregulated transcription of the *c-fos* gene might lead to uncontrolled proliferation²⁶. In view of the possible sequencing of proto-oncogene expression during monocytic differentiation, it would be of interest to determine whether the *c-fms* gene product regulates the expression of *c-myc* or *c-fos* genes. Thus, a decrease in the production of the *c-fms* gene product could lead to defects not only in the control of cell differentiation along the monocyte lineage, but also in cell proliferation.

We thank Drs T. Curran, B. Patterson, J. Cunningham and F. Wong-Staal for providing the *v-fos*, β -actin, *N-ras* and *c-myc* probes, respectively; and Dr J. Griffin for preparation of human monocytes and leukaemic cells. We also thank Ms Eileen Mongeau for secretarial assistance. This work was supported by PHS grant CA34183 from the National Cancer Institute and by an American Cancer Society Faculty Research Award (D.W.K.).

Received 22 February; accepted 14 May 1985.

- McDonough, S. K., Larsen, S., Brodey, R. S., Stock, N. D. & Hardy, W. D. *Jr Cancer Res.* **31**, 953–956 (1971).
- Porzig, K. J., Barbacid, M. & Aaronson, S. A. *Virology* **92**, 91–107 (1979).
- Donner, L., Fedele, L. A., Garon, C. F., Anderson, S. J. & Sherr, C. J. *J. Virol.* **41**, 489–500 (1982).
- Hampe, A., Gobet, M., Sherr, C. J. & Galibert, F. *Proc. natn. Acad. Sci. U.S.A.* **81**, 85–89 (1984).
- Roussel, M. F., Rettenmier, C. W., Look, A. T. & Sherr, C. J. *Molec. cell. Biol.* **4**, 1999–2009 (1984).
- Manger, R., Najita, L., Nichols, E. J., Hakomori, S. & Rohrschneider, L. *Cell* **39**, 327–337 (1984).
- Muller, R. *Cold Spring Harb. Conf. Cell Prolif.* **10**, 451–468 (1983).
- Muller, R. *et al. Molec. cell. Biol.* **3**, 1062–1069 (1983).
- Muller, R., Tremblay, J. M., Adamson, E. D. & Verma, I. M. *Nature* **304**, 454–456 (1983).
- Muller, R., Muller, D. & Guilbert, L. *EMBO J.* **3**, 1887–1890 (1984).
- Bunn, H. F. *et al. Blood* **64**, Suppl. 1, 200a (1984).
- Gonda, T. J. & Metcalf, D. *Nature* **310**, 249–251 (1984).
- Mitchell, R. L., Zokas, L., Schreiber, R. D. & Verma, I. M. *Cell* **40**, 209–217 (1985).
- Slamon, D. J., deKernion, J. B., Verma, I. M. & Cline, M. J. *Science* **224**, 256–262 (1984).
- Groffen, J. *et al. Nucleic Acids Res.* **11**, 6331–6339 (1983).
- van den Berghe, H., Cassiman, J. J., Fryns, D. G., Michaux, J. L. & Sokal, G. *Nature* **251**, 437–438 (1974).
- van den Berghe, H., Michaux, D. G., Sokal, J. L. & Verwilghen, R. *Blood* **48**, 624–626 (1976).
- Collins, S., Gallo, R. & Gallagher, R. *Nature* **270**, 347–349 (1977).
- Westin, E. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 2194–2198 (1982).
- Collins, S., Bodner, A., Ting, R. & Gallo, R. *Int. J. Cancer* **25**, 213–218 (1980).
- Rovera, G., O'Brien, T. G. & Diamond, L. *Science* **204**, 868–870 (1979).
- McCarthy, D. *et al. Leukemia Res.* **7**, 51–55 (1983).
- Major, P., Griffin, J., Minden, M. & Kufe, D. *Leukemia Res.* **5**, 429–430 (1981).
- Grosso, L. & Pitot, H. *Biochem. biophys. Res. Commun.* **119**, 473–480 (1981).
- Kelly, K., Cochran, B. H., Stiles, C. D. & Leder, P. *Cell* **35**, 603–610 (1983).
- Greenberg, M. & Ziff, E. *Nature* **311**, 433–438 (1984).
- Muller, R. & Wagner, E. *Nature* **311**, 438–442 (1984).
- Todd, R. & Schlossman, S. *Blood* **59**, 775–786 (1982).
- Chirgwin, J., Przybyla, A., MacDonald, R. & Rutter, W. *Biochemistry* **18**, 5294–5299 (1979).
- Thomas, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).
- Yam, L., Li, C. & Crosby, W. *Am. J. clin. Path.* **55**, 283–290 (1971).
- Dalla-Favera, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 6497–6501 (1982).
- Murray, M. *et al. Cell* **33**, 749–757 (1983).
- Cleveland, D. W. *et al. Cell* **20**, 95–105 (1980).
- Curran, C., Peters, G., van Beveren, C., Teich, N. M. & Verma, I. M. *J. Virol.* **44**, 674–682 (1982).

Functional expression of a microinjected E_α^d gene in C57BL/6 transgenic mice

Ken-ichi Yamamura*, Hitoshi Kikutani*, Virginia Folsom†, Linda K. Clayton†‡, Masao Kimoto*, Shizuo Akira*, Shin-ichiro Kashiwamura*, Susumu Tonegawa† & Tadimitsu Kishimoto*

* Division of Cellular Immunology, Institute for Molecular and Cellular Biology, and Department of Medicine, Osaka University, 1-3 Yamada-Oka, Suita City, Osaka 565, Japan

† Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The class II major histocompatibility antigens, I-A and I-E, have been detected on the surface of certain immunocompetent cells, including B lymphocytes and monocytes¹⁻³. These molecules are involved in cell-cell interactions in the immune responses^{4,5}. Each class II antigen consists of two subunits, α and β chains, and the genes encoding these subunits have been well characterized at the molecular level⁶⁻⁹. To analyse the regulatory mechanism of E_α gene expression and the role of the I-E antigen in the regulation of the immune responses, we have produced transgenic mice by microinjecting cloned E_α^d genes into fertilized eggs of C57BL/6 mice of *b* haplotype. This strain of mouse carries a deletion in the upstream (5') region of the E_α gene covering the transcriptional promoter and, therefore, does not express this gene¹⁰. Interestingly, this genetic defect of the E_α gene is accompanied by the inability of the host mouse to respond to a certain set of antigens, phenomena generally termed *Ir* gene control^{11,12}. We report here that the E_α^d genes are expressed in these transgenic mice to form the I-E $\alpha^d E_\beta$ antigen on the surface of B lymphocytes and monocytes and that these I-E antigens are functional in terms of the induction of a mixed lymphocyte reaction and the restoration of immune responsiveness to poly(L-glutamic acid-L-lysine-L-phenylalanine) (GL-Phe).

‡ Present address: Dana Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.

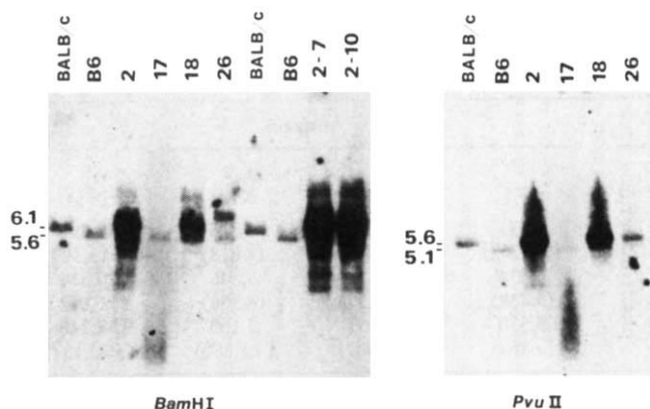


Fig. 1 Southern blot analysis of the E_α gene in transgenic mice³². DNAs (5 μ g) from liver of BALB/c, C57BL/6 or transgenic mice were digested with *Bam*HI or *Pvu*II, electrophoresed on a 0.5% agarose gel, transferred to a nitrocellulose filter and hybridized with the ³²P-labelled 2.6-kb *Sal*I fragment which includes the first exon of the E_α gene. The host C57BL/6 mice carry a deletion in this region³². Mice 2, 17, 18 and 26 were born from ova injected with E_α^d DNA. Mouse 17 did not contain injected genes. The offspring (nos 2-7 and 2-10) of mouse 2 were produced by mating mouse 2 with a female C57BL/6 mouse. DNAs from mice 2-7 and 2-10 were also analysed by *Bam*HI digestion. Sizes of DNA fragments are in kilobases.

Approximately 200 copies of the 14-kilobase (kb) *Sac*II/*Xho*I fragment containing the entire E_α^d gene sequence were microinjected into each fertilized egg of the C57BL/6 mouse. In total, 30 mice were born. These mice were partially hepatectomized and the DNA isolated from the livers was analysed by Southern blotting to determine whether the injected E_α^d gene is retained in the host DNA. Because of the aforementioned deletion, BALB/c DNA yielded a DNA fragment 0.5 kb larger than that of C57BL/6 DNA when these DNAs were digested either with *Bam*HI or *Pvu*II and the digests analysed with a probe covering the 5' end of the E_α gene (see Fig. 1). This allows us to distinguish the injected E_α^d gene from the endogenous defective E_α^b gene present in the host C57BL/6 mouse. As expected, DNA from transgenic mice gave the 5.6-kb *Bam*HI or 5.1-kb *Pvu*II band characteristic of the C57BL/6-derived, defective E_α^b gene (Fig. 1). In addition, DNA from transgenic mice 2 and 18 gave the 6.1-kb *Bam*HI band and 5.6-kb *Pvu*II band diagnostic of the E_α^d gene, indicating that these mice have retained the injected BALB/c E_α^d gene. DNA from the third transgenic mouse, no.

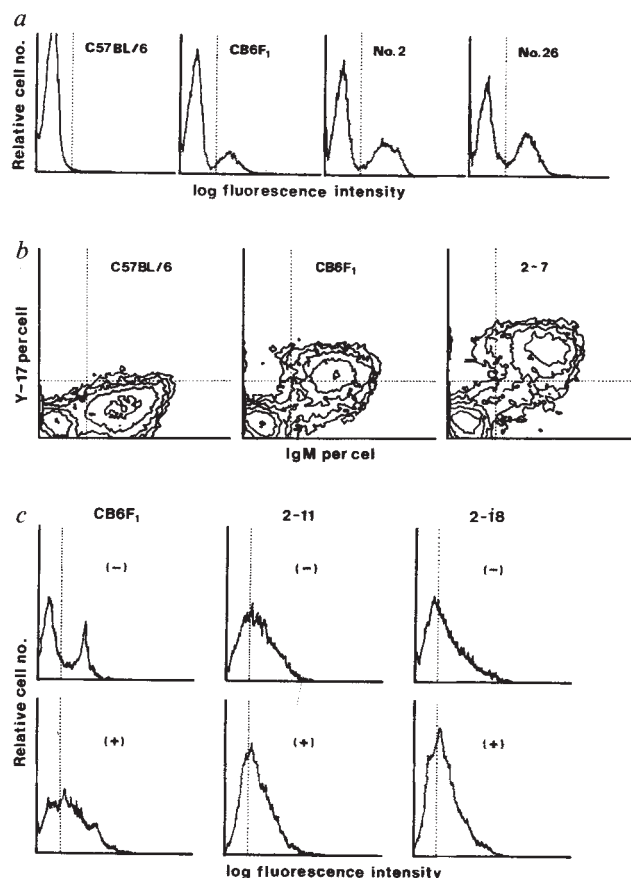


Fig. 2 Expression of the I-E $\alpha^d E_\beta$ antigens on lymphocytes and peritoneal monocytes of the transgenic mice. *a*, Mononuclear cells were prepared from peripheral blood of CB6F₁ and C57BL/6 mice and of mice 2 and 26 on a Ficoll-Hypaque gradient. Cells were incubated with the biotinylated Y17 antibody, followed by fluoresceinated avidin and analysed with a fluorescence-activated cell sorter, FACS-440 (Becton-Dickinson). *b*, Spleen lymphocytes from CB6F₁, C57BL/6 mice and an offspring (no. 2-7) of mouse 2 were incubated with the Y17 antibody and fluorescein isothiocyanate (FITC)-conjugated anti-mouse μ chain, followed by phycoerythrin-conjugated avidin and analysed with a cell sorter. *c*, Peritoneal exudate cells were prepared from CB6F₁ mice and from the offspring of mouse 2 which had been injected intraperitoneally with 1 ml 3% proteose peptone 3 days earlier. Cells (1×10^6 in 1 ml) were incubated in a 24-well plastic plate at 37 °C and non-adherent cells were removed by washing with Hank's balanced salt solution. Adherent cells were cultured with or without 50% (v/v) supernatant of Con A-stimulated spleen cells for 2 days. This supernatant contains, respectively, 100 U ml⁻¹ and 650 U ml⁻¹ γ -interferon and interleukin-2 activity. The cells were then stained with the Y17 antibody and analysed with a cell sorter.

26, also displayed the 5.6-kb *Pvu*II band but the size of the corresponding *Bam*HI fragment was somewhat greater than 6.1 kb. This discrepancy may result from loss of the 5' *Bam*HI site. From an approximate calculation based on the intensities of the E_α^d bands, mice 2 and 18 seem to have at least 10 copies of the E_α^d gene. These E_α^d genes were transmitted to a fraction of offspring, two examples of which (nos 2-7 and 2-10) are shown in Fig. 1. In the following studies, both the original transgenic mice and their offspring were used.

Expression of the E_α gene in the transgenic mice was examined by the surface staining of peripheral blood mononuclear cells and splenic lymphocytes using a monoclonal anti-I-E antibody, Y17 (ref. 13). Because this antibody can recognize combinations of E_α and E_β molecules of various haplotypes, including $E_\alpha^d E_\beta^b$, we could examine whether the product of the microinjected E_α^d gene can pair with the endogenous E_β^b chain and be displayed on the cell surface as the hybrid $E_\alpha^d E_\beta^b$ molecule. As shown in Fig. 2a, about one-third of the peripheral blood lymphocytes from mice 2 and 26 could be stained with Y17, indicating that the lymphocytes of the transgenic mice expressed the hybrid I-E^{d/b} antigen. The fluorescence intensity of the Y17-positive cells of the transgenic mice seems to be significantly higher than that of a CB6F₁ (C57BL/6 × BALB/c F₁) mouse. As expected, the I-E^{d/b} antigen was mainly expressed on B lymphocytes bearing immunoglobulin M (IgM) molecules on the cell surface (Fig. 2b). We also analysed E_α^d gene expression on the monocytes of the transgenic mice. Peritoneal monocytes of the transgenic and CB6F₁ mice, as the control, were cultured for 48 h with or without the culture supernatant of concanavalin A (Con A)-stimulated murine spleen cells containing 50 U ml⁻¹ of γ -interferon and stained with the Y17 antibody. As expected, the I-E^{d/b} molecules were detected on a very small fraction of the unstimulated cells and on most of the stimulated cells of the CB6F₁ mouse. In contrast, most of the peritoneal monocytes from the transgenic mice expressed the I-E^{d/b} antigen in the absence of the Con A supernatant and the addition of the supernatant to the culture media did not increase the level of I-E^{d/b} expression. The reason for this constitutive expression is unknown, but it may be related to the fact that the transgenic mice examined (nos 2-11 and 2-18) carry a large copy number (>20) of E_α^d genes. In any case, it is interesting that not only the injected E_α^d genes but also the endogenous E_β^b gene seem to be constitutively expressed in these mice.

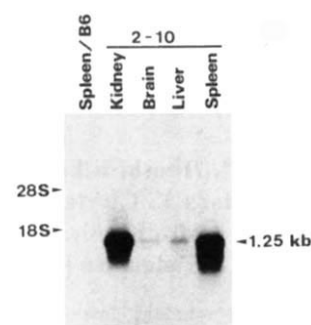


Fig. 3 Northern blot analysis of E_α mRNA from various tissues of a transgenic mouse. Total RNA was phenol-extracted from spleen, liver, kidney and brain of a transgenic mouse and adsorbed to oligo(dT)-cellulose; 2 μ g poly(A)⁺ RNA from spleen and 5 μ g poly(A)⁺ RNA from liver, kidney and brain were denatured, electrophoresed on a 1% formaldehyde-containing agarose gel, transferred to a nitrocellulose filter and hybridized with a ³²P-labelled 3.5-kb *Sal*I fragment containing exons 2, 3 and 4 of the E_α gene as described previously³³. As a control, 2 μ g mRNA from C57BL/6 spleen was also analysed. Sizes of RNA molecules are shown in kilobases.

To examine the transcription of the microinjected E_α^d gene, we performed a Northern blot analysis of the E_α transcript in various tissues from a transgenic mouse, no. 2-10. Figure 3 shows that detectable amounts of E_α messenger RNA of 1.25 kb were observed in poly(A)⁺ RNA from all analysed tissues of the transgenic mouse. The microinjected E_α^d gene seemed to be transcribed at a low level in liver and brain cells. Although kidney cells of a transgenic mouse contained a relatively large amount of E_α mRNA, similar amounts of E_α transcripts were also observed in BALB/c kidneys (data not shown). The class II antigens are known to be expressed on B lymphocytes and macrophages in lymphoid organs^{1-3,14}, liver Kupffer cells^{15,16}, brain astrocytes¹⁶, kidney pelvis cells¹⁵ and blood vessel endothelial cells¹⁷. The observed transcripts of the E_α gene in tissues of the transgenic mouse may be due to these cells. Recently, a rabbit β -globin gene was reported to be expressed in skeletal muscle and testis but not in erythroid cells in transgenic mice¹⁶. Immunoglobulin μ heavy-chain genes were also expressed in heart as well as in lymphocytes¹⁹. We cannot

Table 1 I-E^{d/b} antigen in a transgenic mouse is functional in terms of induction of a mixed lymphocyte reaction and restoration of the response to GL-Phe

	Stimulator cells				Antigens			
	(-)	C57BL/6 (A ^b E ⁻)	Transgenic (A ^b E ^{b/d})	CB6F ₁ (A ^{b/d} E ^{b/d})	(-)	GL-Phe	KLH	PPD
C57BL/6	5,767 (±999)	7,233 (±1,149)	25,622 (±1,459)	90,557 (±15,417)	6,445 (±1,268)	6,326 (±837)	6,373 (±818)	47,296 (±13,307)
Transgenic	—	—	—	—	4,434 (±438)	15,228 (±4,612)	5,846 (±306)	20,314 (±5,061)
CB6F ₁	—	—	—	—	5,530 (±865)	25,363 (±1,067)	5,160 (±1,580)	58,340 (±22,137)

2.5 × 10⁵ responder lymph node cells from normal C57BL/6 mice were stimulated with 1 × 10⁶ mitomycin C (MMC)-treated cells from C57BL/6, CB6F₁ or transgenic mice in 0.2 ml complete culture medium in a microtitre plate (Falcon 3072, Becton-Dickinson) for 4 days. Complete culture medium consisted of RPMI 1640 culture medium, 10% heat inactivated fetal calf serum (FCS), 5 × 10⁻⁵ M 2-mercaptoethanol, 100 mM HEPES, 100 U ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin and 2 mM glutamine. MMC treatment of stimulator cells: 5 × 10⁶ spleen cells were incubated in 1 ml complete culture medium in the presence of 50 μ g MMC at 37 °C in 5% CO₂ in air; after 30 min, cells were washed three times with Hank's balanced salt solution. Cells were pulsed with 1 μ Ci ³H-TdR for the final 16 h of culture. Cells were collected on a filter paper with a semi-automatic cell harvester (Labo Mash Science, Tokyo) and the uptake of ³H-TdR was measured by standard scintillation counting. The proliferative responses of responder cells are expressed as the uptake of ³H-thymidine (mean ± s.d. of triplicate cultures). The antigen-specific T-cell proliferative response was assayed according to the methods of Corradin *et al.*³¹. C57BL/6 mice (three in one group), CB6F₁ mice (three in one group) and a transgenic mouse were immunized with 50 μ g GL-Phe (given by Dr R. Schwartz, NIH) and emulsified in the complete Freund's adjuvant. After 1 week, draining lymph nodes were removed and single-cell suspensions made. Cells were cultured in complete culture medium containing 10% heat-inactivated horse serum instead of FCS in the presence of 200 μ g ml⁻¹ GL-Phe, 100 μ g ml⁻¹ KLH or 50 μ g ml⁻¹ purified protein derivative (PPD) from *Mycobacterium tuberculosis* (Mitui Pharmaceuticals, Tokyo). Cells were cultured for 4 days and the proliferative response assayed as described above.

exclude the possibility that the microinjected E_α^d genes are expressed in inappropriate cells of the transgenic mouse. To assess this possibility more precisely, further studies, including *in situ* RNA hybridization using tissue samples, are required.

Finally, we examined whether the newly expressed I-E^{d/b} antigen has immunological functions. Spleen cells of the transgenic mice could induce a significant proliferation of C57BL/6 lymphocytes (Table 1), suggesting that the I-E^{d/b} antigens of the transgenic mice could be successfully recognized as allo-antigens by the T lymphocytes from C57BL/6 mice. It is known that mouse strains of *H-2b*, *s*, *f* and *q* haplotypes do not respond to certain antigens such as GL-Phe, GL-Leu and pigeon cytochrome *c* (*Ir* gene phenomena)²⁰⁻²²; this unresponsiveness is correlated with the defect in I-E antigen expression in these strains²³. We therefore used responsiveness to GL-Phe as another criterion for the functional expression of the I-E^{d/b} antigens in the transgenic mice, and attempted to demonstrate directly that the *I-E* gene defect is the cause of the unresponsiveness. For this purpose, lymph node cells from mice primed with GL-Phe in complete Freund's adjuvant were assayed for a proliferative response to GL-Phe and, as a control, to keyhole limpet haemocyanin (KLH). We found that GL-Phe could induce significant proliferation of the lymphocytes from CB6F₁ mice (positive control) and from the transgenic mouse but not of lymphocytes from C57BL/6 mice (negative control) (Table 1). The proliferation was specific to the priming antigen (GL-Phe) because no proliferation above the background level was observed in the presence of KLH. These results suggest strongly that the product of the microinjected E_α gene is expressed on the surface of the antigen-presenting cells of the host mouse (C57BL/6) in such a way as to allow the restoration of responsiveness to the antigen GL-Phe.

In the past, several groups have successfully introduced cloned genes for major histocompatibility antigens into cultured cell lines²⁴⁻²⁶, and these transformants have been used to study antigen recognition by T lymphocytes²⁷⁻³⁰. However, this approach is not suitable for *in vivo* analysis of the roles of the major histocompatibility antigens in the generation of the diversity of T-lymphocyte repertoires. To date, this latter type of analysis has been carried out by using H-2 congenic mice, chimaeric mice and H-2 mutant mice. In the present study, we produced transgenic mice which express functional I-E antigens by microinjecting the cloned E_α^d DNA into fertilized eggs of C57BL/6 mice. These transgenic mice have several advantages over the aforementioned mice: (1) the transferred, cloned gene is well characterized; (2) a cloned gene can be modified by *in vitro* mutagenesis or exon shuffling; (3) it is possible to establish new strains carrying the E_α^d gene by crossing these mice with other B6 or B10 congenic mice. Thus, these transgenic mice should be a powerful tool for the analysis of the effect of well-defined class II genes on immune responses.

This study was supported by grants from the Ministry of Education, Science and Culture of Japan (to T.K.), the American Cancer Society (NP-372 to S.T.), US NIH (AI-17879 to S.T.) and the US-Japan Science Cooperation Program sponsored by NSF and JSPS (to S.T. and T.K.).

Received 9 April; accepted 21 May 1985.

- Hammerling, G., Deak, B., Mauve, G., Hammerling, U. & McDevitt, H. O. *Immunogenetics* **1**, 68-74 (1974).
- Steeg, P., Moore, R. & Oppenheim, J. J. *exp. Med.* **152**, 1734-1744 (1980).
- Snyder, D. S., Beller, D. I. & Unanue, E. R. *Nature* **299**, 164-165 (1982).
- McDevitt, H. O. *J. Immunogenet.* **8**, 287-295 (1983).
- Klein, J., Juretic, A., Baxevanis, C. N. & Nagy, Z. A. *Nature* **291**, 455-460 (1981).
- Steinmetz, M. *et al. Nature* **300**, 35-43 (1982).
- Hyldig-Nielsen, J. J. *et al. Nucleic Acids Res.* **11**, 5055-5071 (1983).
- Mathis, D. J., Benoist, C. O., Williams, V. E. II, Kanter, M. R. & McDevitt, H. O. *Cell* **32**, 745-754 (1983).
- Saito, H., Maki, R. A., Clayton, L. K. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5520-5524 (1983).
- Mathis, D. J., Benoist, C., Williams, V. E. II, Kanter, M. R. & McDevitt, H. O. *Proc. natn. Acad. Sci. U.S.A.* **80**, 273-277 (1983).
- Benacerraf, B., Green, I. & Paul, W. E. *Cold Spring Harb. Symp. quant. Biol.* **32**, 569-576 (1967).
- McDevitt, H. O. & Sela, M. *J. exp. Med.* **126**, 969-981 (1967).
- Lerner, E. A. *et al. J. exp. Med.* **152**, 1085-1101 (1980).

- Hammerling, G. J. *Transplant. Rev.* **30**, 64-82 (1976).
- Wiman, K. *et al. Nature* **276**, 711-713 (1978).
- Natali, P. G. *et al. Transplantation* **31**, 75-78 (1981).
- Hirschberg, H., Vergh, O. J. & Thorsby, E. *J. exp. Med.* **152**, 249s-255s (1980).
- Lacy, E., Roberts, S., Evans, E. P., Burtenshaw, M. S. & Costantini, F. D. *Cell* **34**, 343-358 (1983).
- Groschedl, R., Weaver, D., Baltimore, D. & Costantini, F. *Cell* **36**, 647-658 (1984).
- Benacerraf, B. & Dorf, M. E. *Cold Spring Harb. Symp. quant. Biol.* **41**, 465-467 (1976).
- Schwartz, R. H., David, C. S., Sachs, D. H. & Paul, W. E. *J. Immun.* **117**, 531-540 (1976).
- Goodenow, R. S. *et al. Science* **215**, 677-679 (1982).
- Dorf, M. E. & Benacerraf, B. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3671-3675 (1976).
- Solinger, A. M., Ulter, M. E., Malgouy, E. & Schwartz, R. H. *J. exp. Med.* **150**, 830-848 (1979).
- Rabourdin-Combe, C. & Mach, B. *Nature* **303**, 670-674 (1983).
- Mallisen, B., Steinmetz, M., McMillan, M., Pierres, M. & Hood, L. *Nature* **305**, 440-443 (1983).
- Mellor, A. L. *et al. Nature* **298**, 529-534 (1982).
- Margulies, D. H., Evans, G. A., Ozato, K. & Seidman, J. G. *J. Immun.* **130**, 463-470 (1983).
- Germain, R. N., Norcross, M. A. & Margulies, D. H. *Nature* **306**, 190-194 (1983).
- Folsom, V. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 2045-2049 (1984).
- Corradin, G. & Chiller, J. M. *J. exp. Med.* **149**, 436-447 (1979).
- Southern, E. *J. molec. Biol.* **98**, 503 (1975).
- Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).

Neutralization of human T-lymphotropic virus type III by sera of AIDS and AIDS-risk patients

Robin A. Weiss*, Paul R. Clapham*, Rachanee Cheingsong-Popov*, Angus G. Dalgleish*, Christopher A. Carne†, Ian V. D. Weller† & Richard S. Tedder†

* Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London SW3 6JB, UK

† Academic Department of Genito-Urinary Medicine and Virology Section, Department of Medical Microbiology, Middlesex Hospital Medical School, London W1P 7PN, UK

Human T-lymphotropic virus type III (LAV, HTLV-III)^{1,2} is aetiological linked to acquired immune deficiency syndrome (AIDS) and persistent general lymphadenopathy (PGL)²⁻⁷. Specific radioimmunoassays (RIA), enzyme-linked assays, immunofluorescence assays (IFA) and immunoblotting techniques are being used widely to detect serum antibodies to HTLV-III in infected patients and in those at risk of infection⁴⁻⁷. However, these assays do not functionally identify those antibodies that neutralize the infectivity of the virus. We have used three methods of titrating serum neutralizing factors: inhibition of syncytium induction, neutralization of envelope pseudotypes of vesicular stomatitis virus (VSV) and reduction of infectivity of HTLV-III for a cell line permissive to virus replication. We report here that sera from subjects in various disease categories possess only low-level neutralizing activity, even when antibodies to viral membrane antigens are present in high titre. Envelope pseudotypes prepared from four HTLV-III isolates made in three different countries are equally sensitive to neutralization by positive sera, including sera from patients yielding two of the virus isolates.

To examine neutralizing properties of sera containing anti-HTLV-III, we initially determined whether such sera would inhibit the induction of syncytia or would neutralize VSV(HTLV-III) pseudotypes. When uninfected cells bearing HTLV-III receptors are mixed in excess with virus-producing cells, large multinucleated syncytia form within 6 h and are an indicator of viral envelope glycoprotein activity⁸. From previous studies⁹⁻¹² of HTLV-I and HTLV-II, we would expect antibodies binding to the functional epitopes of the viral envelope antigen both to inhibit syncytium formation and to neutralize the infectivity of pseudotypes.

Table 1 shows the mean titres of sera from patients naturally infected with one of the three known HTLV strains. As expected, sera containing antibodies to HTLV-I and HTLV-II specifically inhibited syncytium formation induced by these viruses and also

showed high titres of neutralizing activity specific for the respective VSV pseudotypes. In contrast, sera containing anti-HTLV-III showed negligible neutralizing activity even though the antibody titre for IFA antibody to cell-surface HTLV-III antigens was as high or higher than the IFA titres to HTLV-I and HTLV-II membrane antigens displayed by sera containing antibodies to the respective viruses. None of the 12 sera containing anti-HTLV-III showed significant syncytium inhibiting activity; there was slight inhibition of pseudotype plaques by six of the sera, but none had a neutralizing titre greater than 1:50. When sera containing high-titre neutralizing activity to HTLV-I or HTLV-II were tested for neutralization of VSV(HTLV-III), no cross-reaction was detected.

Because the sera containing anti-HTLV-III were from patients with symptoms of the AIDS-related complex, we sought to determine whether asymptomatic, yet seropositive, individuals have stronger neutralizing activity. The results of VSV(HTLV-III) neutralization tests on 97 positive sera (79 from the United Kingdom⁷ and 18 from Africa¹³) are shown in Table 2, categorized according to the clinical status of the patient, or his or her origin. At 1:10 dilution, only 15% of the sera exhibited neutralization of 80% or more VSV(HTLV-III) plaque-forming units (PFU), and these were mainly British or African patients with PGL. However, when a reduction of 40% PFU was taken as a positive score, then the sera of 69% patients from all clinical groups were found to contain neutralizing factors. The neutralizing activity was completely absorbed by protein A of *Staphylococcus aureus* (data not shown), indicating that it was a property of serum immunoglobulins.

The pseudotype neutralization tests had been performed using the American HTLV-IIIB isolate growing in HT-H9 cells¹⁴ as a source of envelope antigen for the VSV pseudotypes. As genomic polymorphisms have been observed, including variation in the putative *env* gene¹⁵, it seemed plausible that the very weak neutralization observed with British and African sera reflected divergence of envelope antigenicity resulting from the spread of HTLV-III through different host populations. Therefore, we prepared VSV(HTLV-III) pseudotypes with the envelope antigens of four reportedly independent HTLV-III isolates, and titrated neutralizing activity in selected sera including two (Bru and Arg) from patients from whom virus isolates (LAV-1 and CBL-1 respectively) had been made. Serum samples from the British patient Arg were available from before and after the date of biopsy leading to isolation of CBL-1.

Table 3 shows that the sera tested had similar neutralizing titres for VSV pseudotypes with the envelopes of French, British and American HTLV-III isolates. The sera did not have significantly stronger neutralizing activity for their homologous virus isolate. Thus, the relatively low neutralizing titres of sera containing anti-HTLV-III would not seem to be directly attributable to strain variation. When four positive sera were combined, the neutralizing titre was no higher than that of the strongest serum in the pool. Furthermore, an American commercial immunoglobulin G preparation, pooled from several hundred sera selected for hepatitis B virus antibodies and possessing high anti-HTLV-III by RIA, contained barely detectable neutralizing antibodies (Table 3).

It may be argued that the low neutralizing titres of sera containing anti-HTLV-III assayed by syncytium inhibition and by pseudotype inactivation do not reflect the sensitivity to neutralization of HTLV-III itself. We therefore compared the effect on HTLV-III infectivity of four sera previously found to have high IFA titres and relatively strong, weak or undetectable neutralizing activity for VSV(HTLV-III) pseudotypes. Table 4 shows that the neutralizing titres for infectious units of HTLV-III and of VSV(HTLV-III) were closely parallel, although inactivation of HTLV-III may be slightly more sensitive. Low-titre neutralizing activity of human sera to HTLV-III is also being reported by others^{16,17,36,37}. Thus, the rapid and quantitative VSV pseudotype neutralization test serves as an accurate estimate of the protective effect of sera in neutralizing HTLV-III infectivity.

Table 1 Mean titres of human sera from subjects infected with HTLV types I, II or III assayed by different methods

Virus	No. of sera	Competitive RIA	Membrane IFA	Syncytium inhibition	Pseudotype neutralization
HTLV-I	9	1,216	2,250	180	3,257
HTLV-II	4	774	2,500	825	3,750
HTLV-III	12	1,272	2,847	<10	<10

Human sera known to be positive by RIA for antibodies to HTLV-I, -II or -III were titrated after heat inactivation (56°C for 30 min) by methods described previously^{7,10-12}. Values are the geometric means of the reciprocals of the highest serum dilution giving >50% inhibition radioimmunoassay (RIA), >80% fluorescing cells (IFA) and >80% reduction of nuclei in syncytia or of PFU of VSV pseudotype. For both syncytial and pseudotype assays of HTLV-I and HTLV-II, Human osteosarcoma (HOS) cells were used as indicator cells¹¹. JM cells served as indicator cells for syncytial assays of HTLV-III⁸; adherent CEM cells pretreated with 25 µg ml⁻¹ DEAE-dextran were used for assay of VSV(HTLV-III) pseudotypes, and were subsequently overlaid with HOS cells and then agar medium, as described previously⁸.

Table 2 Neutralization of VSV(HTLV-III) pseudotype by sera of subjects positive by RIA

Serum donor	No. of sera	No. of sera reducing VSV(HTLV-III) plaques by	
		>40%	>80%
AIDS, KS	6	6	1
AIDS, OI	6	2	0
PGL only	13	12	4
PGL plus minor OI	4	4	2
Contacts of AIDS or PGL	9	7	2
Symptomatic homosexuals	8	6	1
Asymptomatic homosexuals	11	6	0
Haemophiliacs	22	11	0
Africans with KS or PGL	18	13	4
Total seropositives	97	67	14
Seronegative controls	20	0	0

Heat-inactivated sera were incubated at 1:10 dilution with 5×10^2 PFU VSV(HTLV-IIIB) for 1 h at 33°C and surviving PFU were plated on adherent CEM cells as described for Table 1. Positive sera taken in the United Kingdom were selected from a previous study⁷, as were the uninfected control sera. Africans with Kaposi's sarcoma (KS) or PGL were Zambians¹³. Patients with AIDS were identified clinically as opportunistic infection (OI) and/or KS according to the US Centers for Disease Control definition. Patients with PGL had enlarged lymph nodes (>1 cm) in two or more extra-inguinal sites persisting for more than 3 months according to the Centers for Disease Control definition. Two patients had oral candidiasis, one developed a lung abscess following aspiration during an epileptic fit. One patient developed AIDS 12 months after the serum specimen was taken. AIDS and PGL contacts were the regular sexual partners for at least 6 months before diagnosis of AIDS or PGL in their partner. All were well and without lymphadenopathy. Symptomatic homosexuals included six patients with recent viral-like illnesses accompanied by lymphadenopathy (4), pharyngitis (2) and an unexplained transient erythematous macular rash (1), one patient with generalized lymphadenopathy and malaise of <3 months' duration and one patient with a streptococcal pneumonia. Asymptomatic homosexuals were men attending the clinic with no systemic symptoms or generalized lymphadenopathy. Haemophiliacs had received regular clotting factor replacement therapy. Neutralization tests were performed and scored 'blind', that is, without knowledge of RIA titres or clinical status of the patient.

The low titres of virus neutralizing antibodies in subjects infected with HTLV-III contrasts with the type-specific, high-titre human humoral response to HTLV-I and HTLV-II⁹⁻¹², and also differs from the immune response of animals naturally infected with avian¹⁸, feline¹⁹ and bovine²⁰ leukaemia viruses. HTLV-III is morphologically more closely related to the lentivirus family of retroviruses, reportedly exhibiting weak cross-reaction of antibody to the core antigen of equine infectious anaemia virus (EIAV)²¹, and partial genomic homology to visna virus (VV)²². Strongly neutralizing sera are commonly detected in animals infected with EIAV^{23,24} and VV^{25,26}, although neutralizing activity has not been found in goats infected with

Table 3 Neutralization of VSV pseudotypes with envelopes of different HTLV-III isolates including matched sera

Serum donor	Titre neutralizing VSV(HTLV-III) plaques (%)							
	LAV-1 (France)		CBL-1 (UK)		IIIB (USA)		IIIA (USA)	
	40	80	40	80	40	80	40	80
Bru (LAV-1)	50	10	50	10	50	10	50	10
Arg (CBL-1) October 1983	50	<10	25	<10	50	<10	50	<10
Arg (CBL-1) September 1984	50	<10	50	<10	50	<10	50	<10
Neb	50	10	50	10	50	10	50	10
Sel	250	50	250	50	250	50	250	50
Pooled IgG (USA)	50	<10	50	<10	50	<10	50	<10
Uninfected control	<10	<10	<10	<10	<10	<10	<10	<10

VSV pseudotypes were prepared with four isolates of HTLV-III and plated on adherent CEM cells (Table 1) following incubation with the sera indicated. The CBL-1 isolate was made from a lymphomatous lymph node biopsy taken from AIDS patient Arg in June 1984, 8 months after the first serum sample and 3 months before the second serum sample assayed (A.G.D., A. Horwich, M. G. Peckham and R.A.W., unpublished data); this patient was a homosexual living in the United Kingdom with a single sexual partner who was also seropositive. Serum Bru was taken from a PGL patient at the time of virus (LAV-1) isolation¹. Fivefold serial dilutions of sera were made and the reciprocal of the highest dilution causing >40 and >80% reduction of pseudotype PFU are tabulated.

Table 4 Neutralization of HTLV-III infectivity by selected human sera

Donor	Serum Membrane IFA titre	Surviving infectious units after antiserum treatment	
		HTLV-III	VSV(HTLV-III)
Control	0	5,000	5,000
Arb	1,250	500	500
Neb	2,500	50	500
Sel	2,500	<50	<10
Dol	1,250	5,000	5,000

Neutralization of infectivity of HTLV-III was measured by end-point titration of surviving virus after adding 1:10 dilution of antiserum to 5,000 infectious units of virus stock (cell-free culture supernatant) for 1 h at 33 °C. Duplicate flasks of HT-H9 cells¹⁴ were infected with serial 10-fold dilutions of the treated virus in the presence of 25 µg ml⁻¹ DEAE-dextran. The cells were maintained and passaged in growth medium and the percentage of cells immunofluorescent for HTLV-III membrane antigens using a known human serum assayed on days 7, 10, 13 and 18 post-infection. The cells in each flask were also tested for virus production by syncytium-inducing activity in JM cells, with similar or slightly more sensitive results than IFA. The end-point titre at 18 days (taking >10% fluorescent cells as positive) was used to calculate the surviving units of HTLV-III infecting the HT-H9 cells following antiserum treatment. The surviving plaques of VSV(HTLV-III) after similar serum treatment of 5,000 pseudotype PFU were assayed on adherent CEM cells as described in Table 1 legend.

caprine arthritis encephalitis lentivirus (CAEV)²⁷. Considerable antigenic variation is observed with EIAV and VV; antibodies elicited early in infection of individual animals have a narrow specificity of neutralizing activity, allowing non-neutralizable virus variants to emerge²⁵⁻²⁶, whereas later sera can neutralize a broader range of virus variants²⁵.

Antigenic variation might explain the apparently low level of neutralization of HTLV-III by human sera. However, our preliminary data show that neither early nor late serum relative to HTLV-III isolation from one patient was strongly neutralizing, and that different virus isolates supposedly made in France, the United States and the United Kingdom did not vary significantly in sensitivity to neutralization by selected sera. Preliminary experiments using human or guinea pig complement indicate that complement does not markedly enhance the inactivation of VSV(HTLV-III) pseudotypes by weakly neutralizing sera, again reminiscent of CAEV (ref. 27). This raises the question of the nature of the membrane antigens recognized by human sera on the surface of HTLV-III-infected cells, which is under investigation using biochemical and immunoelectron microscope techniques.

The CD4 (T₄) cell-surface antigen is an essential and specific component of the HTLV-III receptor^{8,28}. Because neutralizing antibodies generally bind to those epitopes of retrovirus envelope glycoproteins that interact with receptors^{29,30}, the low

neutralizing response might be explained by molecular mimicry between the HTLV-III glycoprotein epitopes and an endogenous cellular antigen that normally interacts with CD4. Class II major histocompatibility complex (MHC) antigens (for example, HLA-DR) expressed on activated B cells and on antigen-presenting cells are thought to interact with CD4 antigen during the helper immune functions of CD4-positive cells^{31,32}, but we have not yet observed any anti-MHC class II antibodies that neutralize VSV(HTLV-III) pseudotypes.

Although we found no clear pattern of neutralization response among nearly 100 infected subjects, a larger proportion of PGL patients had detectable neutralizing sera than either AIDS patients or healthy seropositive individuals. The low neutralization titres may explain why patients seropositive by other techniques behave as infective carriers through sexual contact³³ or blood transmission^{34,35}. It is important to understand more about the parameters of HTLV-III neutralization for the development of effective vaccines. It remains to be seen whether HTLV-III envelope antigen manufactured by recombinant DNA techniques will yield a more powerfully protective immunogen than is apparent in humans exposed to live virus.

We thank M. Exley and A. Smith for technical assistance; members of the Middlesex/University College Hospitals AIDS Research Group for sera from carefully categorized homosexual patients; S. Ball, L. Luzzatto and P. Kernoff for haemophilic sera; A. C. Bayley and R. G. Downing for Zambian PGL and KS sera; M. Popovic, R. C. Gallo, L. Montagnier and J.-C. Chermann for American and French HTLV-III isolates; and L. Montagnier for serum Bru. This study was supported by the Wellcome Trust, MRC and the Cancer Research Campaign.

Received 13 February; accepted 3 April 1985.

1. Barré-Sinoussi, F. *et al. Science* **220**, 868-870 (1983).
2. Gallo, R. C. *et al. Science* **224**, 500-503 (1984).
3. Levy, J. A. *et al. Science* **225**, 840-842 (1984).
4. Brun-Vézinet, F. *et al. Lancet* **i**, 1253-1256 (1984).
5. Schupbach, J. *et al. Science* **224**, 503-506 (1984).
6. Safai, B. *et al. Lancet* **i**, 1438-1440 (1984).
7. Cheingsong-Popov, R. *et al. Lancet* **ii**, 477-480 (1984).
8. Dalgleish, A. G. *et al. Nature* **312**, 763-767 (1984).
9. Hoshino, H., Shimoyama, M., Miwa, M. & Sugimura, T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7337-7341 (1983).
10. Nagy, K., Clapham, P., Cheingsong-Popov, R. & Weiss, R. A. *Int. J. Cancer* **32**, 321-328 (1983).
11. Clapham, P., Nagy, K. & Weiss, R. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3083-3086 (1984).
12. Tedder, R. *et al. Lancet* **ii**, 125-128 (1984).
13. Bayley, A. C. *et al. Lancet* **i**, 359-361 (1985).
14. Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. *Science* **224**, 497-500 (1984).
15. Shaw, G. M. *et al. Science* **226**, 1165-1171 (1984).
16. McDougall, J. S. *et al. J. Immun. Meth.* **76**, 171-183 (1985).
17. Robert-Guroff, M., Brown, M. & Gallo, R. C. *Nature* **316**, 72-74 (1985).
18. Weiss, R. A. & Biggs, P. M. *J. natn. Cancer Inst.* **49**, 1713-1725 (1972).
19. Jarrett, O. *Cold Spring Harb. Conf. Cell Prolif.* **8**, 603-612 (1981).
20. Bruck, C., Portetelle, D., Mammerickx, M., Mathot, S. & Burny, A. *Leukemia Res.* **8**, 315-321 (1984).
21. Montagnier, L. *et al. Ann. Virol.* **135E**, 119-134 (1984).
22. Gonda, M. A. *et al. Science* **227**, 173-177 (1985).
23. Kono, Y., Kobayashi, K. & Fukunaga, Y. *Archs Virol.* **41**, 1-10 (1973).
24. Montelaro, R., Parekh, B., Issel, C. & Orrego, A. *J. biol. Chem.* **259**, 1053-1059 (1984).
25. Narayan, O., Griffin, D. E. & Clements, J. E. *J. gen. Virol.* **41**, 343-352 (1978).

26. Scott, J. V., Stowring, L., Haase, A. T., Narayan, O. & Vigne, R. *Cell* **18**, 321-327 (1979).
27. Klevjer-Anderson, P. & McGuire, T. C. *Infect. Immun.* **38**, 455-461 (1982).
28. Klatzmann, D. *et al.* *Nature* **312**, 767-768 (1984).
29. Weiss, R. A. in *Virus Receptors* (eds Longberg-Holm, K. & Philipson, L.) 186-202 (Chapman & Hall, London, 1981).
30. Weiss, R. A., Clapham, P. R., Nagy, K. & Hoshino, H. *Curr. Topics Microbiol. Immun.* **115**, 235-246 (1985).
31. Meuer, S. C., Schlossman, S. F. & Reinherz, E. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4395-4399 (1982).
32. Biddison, W. E. *et al.* *J. exp. Med.* **159**, 783-797 (1984).
33. Gazzard, B. G. *et al.* *Lancet* **ii**, 480-483 (1984).
34. Anonymous report *Lancet* **ii**, 1376-1377 (1984).
35. Weiss, R. A. in *RNA Tumor Viruses* Vol. 2 (eds Weiss, R. A., Teich, N. M., Varmus, H. E. & Coffin, J.) 405-485 (Cold Spring Harbor Laboratory, New York, 1985).
36. Ho, D. D., Rota, T. R. & Hirsch, M. S. *New Engl. J. Med.* **312**, 649-650 (1985).
37. Clavel, F., Klatzmann, D. & Montagnier, L. *Lancet* **ii**, 879-880 (1985).

HTLV-III-neutralizing antibodies in patients with AIDS and AIDS-related complex

Marjorie Robert-Guroff, Monica Brown & Robert C. Gallo

Laboratory of Tumor Cell Biology, National Cancer Institute, Bethesda, Maryland 20205, USA

The isolation of the human T-cell leukaemia (lymphotropic) virus type III (HTLV-III or lymphadenopathy-associated virus) from cells of many patients with acquired immune deficiency syndrome (AIDS)^{1,2} presented the first evidence that the virus was the aetiological agent of the disease. Subsequent seroepidemiological studies have shown the presence of HTLV-III-specific antibodies in the serum of most patients with AIDS and AIDS-related complex (ARC)^{3,4}, and in the serum of many individuals at risk for AIDS³⁻¹⁴. Despite these extensive studies, there are no reports of protective effects of HTLV-III antibodies. In contrast, neutralizing antibodies specific for HTLV-I and -II have been identified previously¹⁵. Therefore, we investigated whether HTLV-III-exposed individuals possess antibody activities capable of inhibiting viral infection (see also refs 16, 17). Here, we report that natural antibodies capable of neutralizing HTLV-III infection of H9 cells were detected in most adult AIDS and ARC patients but in no normal healthy heterosexual controls. Geometric mean antibody titres in ARC patients were double those in AIDS patients, and were even higher in two antibody-positive healthy homosexuals. This suggests that virus neutralizing antibodies may exert an *in vivo* protective effect. The presence of these antibodies indicates an immunological response to HTLV-III which potentially may be manipulated for therapeutic advantage. The methodology used here will be useful in monitoring future vaccine approaches.

We used the H9 clone of the HT cell line¹ as a target for cell-free HTLV-III infection and initially analysed several sera for virus neutralizing activity. Infection of H9 cells was assessed by monitoring expression of HTLV-III p24 using a monoclonal antibody¹⁸ in an indirect immunofluorescence assay. Figure 1a illustrates the kinetics of infection of H9 cells with HTLV-III virus preincubated with sera positive or negative for virus neutralizing activity. By 3 days post-infection, ~80% of the H9 cells incubated with HTLV-III pretreated with control serum were infected, as indicated by their expression of HTLV-III p24. In contrast, only 10% of H9 cells expressed HTLV-III p24 at day 3 when infected with virus pretreated with serum from a patient with ARC. That inhibition of infection was mediated by a viral rather than a cellular antigen was suggested by ready infection of H9 cells with HTLV-III following pretreatment of the cells rather than the virus with the same sera (Fig. 1a). The inhibitory activity of certain sera was not simply a nonspecific effect of high serum concentrations because the activity was titratable (Fig. 1b).

We also detected neutralizing activity by other criteria. For example, two serum dilutions giving 95 and 90% neutralization

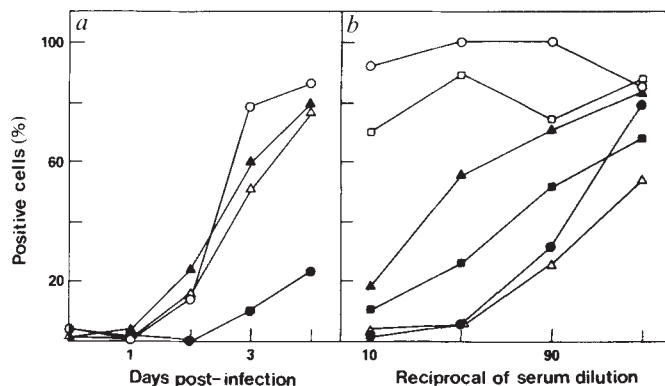


Fig. 1 *a*, Kinetics of infection of H9 cells in the presence of human serum. Sera from a normal healthy heterosexual (○) and from a patient with ARC (●). In a parallel experiment, the same sera were preincubated with H9 cells for 1 h at 4 °C. Cells were washed with PBS, incubated with the cell-free HTLV-III preparation, and cultured as described below. Results of this treatment of the H9 cells with sera are represented by the normal, healthy heterosexual serum (△) and the serum of the patient with ARC (▲). *b*, Titrations of representative human sera for virus neutralizing antibody activity. Representative titrations are shown for two sera negative for viral neutralizing antibody activity: a patient with AIDS (○) and a patient with ARC (□). Representative sera positive for virus neutralizing antibody were obtained from a paediatric AIDS case (●), a patient with ARC (■), a healthy homosexual (△) and an adult AIDS patient (▲). All values obtained were normalized to the level of infection attained in the presence of a standard antibody-negative serum.

Methods. *a*, Media containing 2.5×10^8 HTLV-III particles per ml were collected from H9/HTLV-III cells. Following removal of the cells by low-speed centrifugation, the virus-containing supernatant was centrifuged for 3 h at 32,000g in a Beckman type 35 rotor. Viral pellets were resuspended in a total volume of 2.25 ml media (RPMI 1640 containing 20% fetal calf serum and penicillin/streptomycin). Uninfected H9 cells were washed in media and incubated for 20 min at room temperature in media containing $2 \mu\text{g ml}^{-1}$ polybrene. Cells were washed and resuspended at a concentration of $4 \times 10^6 \text{ ml}^{-1}$. Sera to be tested were heat inactivated at 56 °C for 30 min and filter sterilized. For each assay, 20 μl virus suspension and 20 μl of a 1:2 dilution of serum were mixed and incubated in a microtitre plate for 1 h at 4 °C and 15 min at room temperature. H9 cells (10 μl) were added to each well and incubation continued for 1 h at 37 °C. Aliquots (15 μl) of each mixture were plated into 200 μl media in duplicate wells of another microtitre plate. Cultures were incubated at 37 °C in an incubator in which the CO_2 concentration was maintained at 5%. At the indicated time periods, cultures in individual wells were removed, washed twice with PBS and once with 1:1 PBS/water. Cells were suspended in ~30 μl of the same solution and 5–10- μl aliquots were spotted on eight-well toxoplasmosis slides. The spots were air-dried and the cells fixed for 10 min in 1:1 methanol/acetone at room temperature. Slides were stored at -20 °C before the assay. Viral infection was monitored by indirect immunofluorescence assay as described previously²⁴ on the fixed cells using ascites fluid containing monoclonal antibody BT3 (Biotech Research Laboratories) specific for HTLV-III p24 at a dilution of 1:200. Ascites fluid derived from mice inoculated with P3 $\times$ 63 cells was used as a negative control at a similar dilution. A fluorescein isothiocyanate conjugated F(ab')₂ fragment of sheep anti-mouse IgG was used as second antibody. Per cent fluorescent cells was determined following counts of at least 100 cells on each duplicate preparation. *b*, Sera were diluted in media as indicated and assayed as for *a*. Viral infection, monitored by expression of HTLV-III p24 in the indirect immunofluorescent assay, was determined at 3 days post-infection.

when monitored using monoclonal anti-p24 gave 89% and 95% inhibition, respectively, when monitored using a specific monoclonal antibody to HTLV-III p15 (ref. 18). Measurement of reverse transcriptase activity in proportionately increased cultures showed that the same samples neutralized viral infectivity on day 3 by 62% and 51% and on day 7 by 76% and 54%. Our system was optimized for detection of viral p24 expression in

Table 1 HTLV-III neutralizing activity is a property of IgG and is directed against a viral antigen

Serum sample	Patient diagnosis	Serum treatment	HTLV-III neutralizing antibody titre*
Experiments with purified IgG			
1	ARC	None	115
		Purification of IgG	110
2	AIDS	None	34
		Purification of IgG	40
3	ARC	None	60
		Purification of IgG	76
Absorption experiments			
4	ARC	None	135
		Absorbed with HTLV-III	<10
		Absorbed with H9 cells	120
		Absorbed with H9/HTLV-III	90
5	AIDS	None	75
		Absorbed with HTLV-III	<10
		Absorbed with H9 cells	22
		Absorbed with H9/HTLV-III	34
6	ARC	None	>270
		Absorbed with HTLV-III	50
		Absorbed with H9 cells	>270
		Absorbed with H9/HTLV-III	>270

IgG was purified from 0.5-ml aliquots of human serum by absorption to protein A-Sepharose equilibrated in phosphate-buffered saline (PBS). Following extensive washing of the columns with PBS, IgG was eluted with 0.1 M glycine-HCl, pH 2.8. The eluate was neutralized with 2 M Tris-HCl, pH 8.0, dialysed extensively against 10 mM ammonium bicarbonate and lyophilized. The purified fractions were dissolved in 0.5 ml PBS, filter sterilized and diluted in media for titration of neutralizing antibody activity as described in Fig. 1 legend. For virus absorption experiments, 62 ml cell-free virus supernatant containing $2-5 \times 10^8$ virus particles per ml were pelleted as described in Fig. 1 legend. The viral pellet was resuspended in 100 μ l of a 1:10 dilution of serum to be absorbed and incubated overnight at 4 °C. The virus was again pelleted by centrifugation and the absorbed serum was saved for titration of virus neutralizing antibody activity. Sera were similarly absorbed on pellets of washed 10^7 cells and titrated.

* Values for % HTLV-III p24-positive cells were normalized to the level of infection obtained in the presence of a standard negative serum treated similarly to the test serum. Antibody titres were then expressed as the reciprocal of the serum dilution at which virus infection was 60% of that obtained in the presence of this standard negative serum.

cells washed free of extracellular virus. The enzyme levels in culture fluids reflected not only activity in newly released virus, but also residual activity from the large virus inoculum. Ideally, precise quantitation of neutralization by enzyme assay would require a limiting inoculum and assay after 10–14 days of culture. Nevertheless, the decrease in reverse transcriptase activity obtained substantiated our hypothesis that positive sera blocked infectivity and did not simply alter p24 detection. In subsequent experiments, we assessed infectivity by p24 expression.

Serum immunoglobulins were purified on protein A-Sepharose columns and assayed for virus neutralizing activity. The serum activity capable of inhibiting viral infection resided in the IgG fractions (Table 1), indicating that the activity is immunoglobulin mediated. It is unlikely that these fractions are contaminated extensively with viral envelope capable of competing for viral receptors, or with some other non-immunoglobulin factor, but we have not eliminated these possibilities. To confirm that the inhibitory activity was directed against a viral, rather than a cellular, antigen, sera were absorbed with preparations of cell-free virus or with infected or uninfected H9 cells. Although absorption with cells had little effect, absorption with viral preparations effectively removed the inhibitory activity of positive sera (Table 1). The fact that similar titres were obtained following absorption with infected and uninfected H9 cells suggests that viral antigenic determinants involved in the neutralizing antibody response are masked or presented in a

Table 2 HTLV-III neutralizing antibody in AIDS, ARC and other at-risk patients

Serum source	HTLV-III-neutralizing antibody			Geometric mean titre $\pm$ s.e.m.
	No. positive/ no. tested	% Positive	Range of titre	
Adult AIDS patients	21/35	60	10–520	44 $\pm$ 10
Paediatric AIDS patients	3/9	33	80–180	117 $\pm$ 28
Adult ARC patients	28/35	80	17–560	88 $\pm$ 16
Healthy homosexuals*	2/12	17	130–340	210 $\pm$ 101
Healthy heterosexuals	0/20	0	—	—
Heterosexual partners of AIDS patients†	1/3	33	78	—
Mothers of paediatric AIDS patients‡	0/2	0	—	—
Siblings of AIDS patients	1/2	50	55	—
Patients with acute mononucleosis	1/4	25	13	—
Patient with sarcoidosis	0/1	0	—	—

All sera were screened for virus neutralizing antibody at a 1:10 dilution. Those sera possessing activity were further titrated as described in Fig. 1 legend. Antibody titre is defined in Table 1 legend.

* Two of the 12 sera were also positive for HTLV-III antibodies by enzyme-linked immunosorbent assay and Western blot assay³. A third serum showed weak reactivity with p24.

† All three individuals were positive for HTLV-III antibodies.

‡ Sera from these two foster mothers were negative for HTLV-III antibodies.

|| The positive sibling was also positive for HTLV-III antibody.

different conformation on virus-infected cells compared with mature virions.

Having defined a system able to detect HTLV-III neutralizing antibodies, we analysed sera from additional patients (Table 2). A high prevalence of patients with either AIDS or ARC possess virus neutralizing antibodies, in contrast to healthy heterosexual individuals in whom no such activity was demonstrated. Although neutralizing antibody titres ranged upwards of 500 for both groups of patients, overall titres were low, but ARC patients possessed a twofold higher geometric mean antibody titre compared with that of the AIDS patients. Moreover, among healthy homosexuals at risk for AIDS, the geometric mean antibody titre (determined on only two antibody-positive individuals, and therefore not appropriate for statistical analysis) was higher than that of either of the two patient groups. Whether this trend of higher titre with less disease manifestation will be substantiated, whether it suggests a protective effect of the neutralizing antibodies, or whether it signifies a general decline in antibody titre with progression of the disease requires further investigation.

Although the number of cases examined was small, paediatric AIDS patients positive for neutralizing antibodies had elevated titres compared with those of adult AIDS patients. This difference may reflect different criteria used in diagnosing paediatric and adult AIDS cases. Among miscellaneous cases, we detected neutralizing antibodies in the serum of a heterosexual partner of an AIDS patient and in the serum of a sibling of an AIDS patient. HTLV-III neutralizing antibody activity was not detected in any healthy heterosexual individuals (Table 2). However, a very low titre was obtained in one of four serum samples from patients with acute mononucleosis. This result suggests some weak cross-reactivity with viral antigen in sera possessing high levels of heterophilic antibodies.

In other retroviral systems, the major envelope glycoprotein is the target for neutralizing antibody^{19,20}. We anticipate this will also be the case here, and studies are under way to define the antigen recognized. Whether these virus neutralizing antibodies lead to *in vivo* protection, and whether they will be relevant to future vaccine approaches, are yet to be determined. Knowledge of other retroviruses leads to the question of whether the role of HTLV-III neutralizing antibody is similar to that in the visna virus system, where infection is persistent and the resultant disease is slowly progressive. Neutralizing antibodies elicited by visna virus have a narrow range of specificity and exert a selective pressure, leading to replication of non-neutralized mutant viruses during the course of the disease²¹. This

mechanism may be relevant to AIDS because HTLV-III isolates have shown a range of genomic variability, especially in the envelope region²², and a relatedness of HTLV-III to visna virus was demonstrated recently²³. It will be necessary to obtain serial isolates and serum samples from individuals throughout the course of the disease to determine whether HTLV-III neutralizing antibodies exert selective pressure on viral mutants.

The presence of HTLV-III neutralizing antibodies in sera of individuals exposed to HTLV-III demonstrates an immunological response during the course of disease development which may be used therapeutically. Furthermore, it gives hope that appropriate vaccine approaches may be effective in preventing viral infection. Further studies will determine whether virus neutralizing antibodies in the sera of patients have prognostic value or will be indicative of appropriate treatment regimes.

We thank Dr M. V. O'Shaughnessy for helpful discussion, Dr R. E. Tarone for statistical analysis and Ms Anna Mazzuca for assistance in preparing the manuscript.

Received 28 January; accepted 21 May 1985.

1. Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. *Science* **224**, 497-500 (1984).
2. Gallo, R. C. *et al.* *Science* **224**, 500-502 (1984).
3. Sarngadharan, M. G., Popovic, M., Bruch, L., Schupbach, J. & Gallo, R. C. *Science* **224**, 506-508 (1985).
4. Safai, B. *et al.* *Lancet* **i**, 1438-1440 (1984).
5. Goedert, J. J. *et al.* *Lancet* **ii**, 711-716 (1984).
6. Tsoukas, C. *et al.* *New Engl. J. Med.* **311**, 1514-1515 (1984).
7. Goedert, J. J. *et al.* *Blood* **65**, 492-495 (1985).
8. Gazzolo, L., Gessain, A., Robin, Y., Robert-Guroff, M. & de-The, G. *New Engl. J. Med.* **311**, 1252-1253 (1984).
9. Redfield, R. R. *et al.* *J. Am. med. Ass.* **253**, 1571-1573 (1985).
10. Harris, C. A. *et al.* (in preparation).
11. Clumeck, N. *et al.* *New Engl. J. Med.* **310**, 492-497 (1984).
12. Saxinger, W. C. *et al.* *Science* **227**, 1036-1039 (1985).
13. Clumeck, N. *et al.* (in preparation).
14. Gazzolo, L. *et al.* *Int. J. Cancer* (in the press).
15. Clapham, P., Nagy, K. & Weiss, R. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2886-2889 (1984).
16. Weiss, R. A. *et al.* *Nature* **316**, 69-72 (1985).
17. McDougal, J. S. *et al.* *J. immun. Meth.* **76**, 171-183 (1985).
18. Veronese di Marzo, F. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
19. Steeves, R. A., Strand, M. & August, J. T. *J. Virol.* **14**, 187-189 (1974).
20. Ihle, J. N. & Lazar, B. J. *J. Virol.* **21**, 974-980 (1977).
21. Narayan, O., Clements, J. E., Griffin, D. E. & Wolinsky, J. S. *Infect. Immun.* **32**, 1045-1050 (1981).
22. Shaw, G. M. *et al.* *Science* **226**, 1165-1171 (1984).
23. Gonda, M. A. *et al.* *Science* **227**, 173-177 (1985).
24. Robert-Guroff, M., Ruscetti, F. W., Posner, L. E., Poiesz, B. J. & Gallo, R. C. *J. exp. Med.* **154**, 1957-1964 (1981).

An anti-idiotypic vaccine against experimental schistosomiasis

J. M. Grzych, M. Capron, P. H. Lambert*, C. Dissous, S. Torres & A. Capron

Centre d'Immunologie et de Biologie Parasitaire, Unité Mixte INSERM U 167-CNRS 624, Institut Pasteur, BP 245, 59019 Lille Cedex, France

* Centre OMS de Recherche et de Formation en Immunologie, Département de Pathologie CMV, 1211 Genève 4, Switzerland

Schistosomiasis is a parasitic infection of man which is widespread in tropical countries, and which so far has resisted attempts at control. We have been approaching the problem from an immunological angle. We have previously reported¹ the production of a rat monoclonal IgG2a antibody against *Schistosoma mansoni* which exhibits marked cytotoxicity for schistosomula in the presence of eosinophils and a high degree of protection by passive transfer in naive rats. This antibody, IPLSm1, was shown to bind specifically to a schistosomulum membrane target antigen defined as a glycoprotein of relative molecular mass 38,000 (38K)², which is strongly immunogenic in schistosomula infection of various animal species including man³. Although theoretically the 38K protein represents an excellent candidate for a potential vaccine against schistosomiasis, the glycanic nature of the epitope recognized by

IPLSm1 limits its production by DNA recombinant technology. It was, moreover, shown that, together with protective antibodies, the 38K molecule was able to induce the production of blocking IgG2c antibodies that inhibit the functional properties of IPLSm1 both *in vitro* and *in vivo*⁴. Therefore, following Jerne's network theory⁵, we considered an alternative approach, the possibility of immunization using anti-idiotypic antibodies. In the present study, rat monoclonal anti-idiotypic antibodies were produced against IPLSm1 (AB₁). Anti-idiotypic antibodies (AB₂) were selected by their capacity to inhibit the binding of radioiodinated AB₁ to its 38K target antigen. Sera from naive LOU rats immunized with a purified AB₂ preparation contained specific anti-schistosome antibodies (AB₃) which bound to 38K. AB₃ antibodies were strongly cytotoxic for schistosomula in the presence of rat eosinophils and conferred highly significant protection by passive transfer. Most importantly, rats immunized with AB₂ demonstrated marked protection (50-80%) to a challenge infection.

Monoclonal antibodies to IPLSm1 were obtained in a homologous hybridization system. Male LOU rats were injected subcutaneously with 1 mg of purified IPLSm1 plus 1 mg of glutaraldehyde-aggregated human IgG in the presence of complete Freund's adjuvant. Two weeks later, the rats received a second injection of the same preparation. The immunization was completed by two subcutaneous injections of 1 mg of

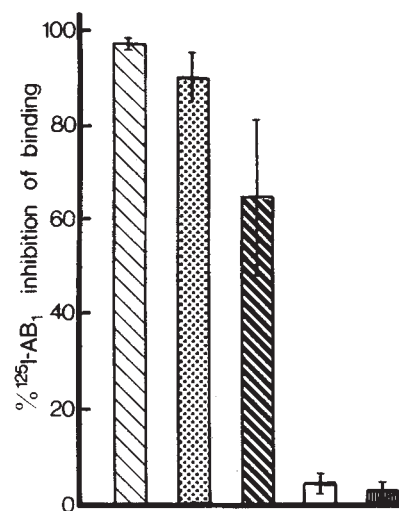


Fig. 1 Inhibition of radiolabelled AB₁ antibody (IPLSm1 IgG2a antibody) to the 38K antigen. Experiments were performed on PVC microtitre plates precoated with C₃-109 IgM antibody (an *S. mansoni* rat monoclonal antibody which recognizes the 38K molecule but does not interfere with IPLSm1 binding). Each well of the PVC plates was coated with 100 µl of a 10 µg ml⁻¹ solution of C₃-109 IgM diluted with 10 mM phosphate-buffered saline (PBS), a procedure which gives better fixation of the 38K antigen. After 2 h, the plates were washed twice in PBS buffer and saturated for 30 min with 200 µl of a 2% bovine serum albumin (BSA) solution in PBS. The plates were then washed twice in PBS-0.1% BSA. A Nonidet P-40 extract of schistosomula was added (100 µl antigen solution containing 100 µg of protein). After 2 h exposure at 37 °C, plates were washed twice in PBS-0.1% BSA buffer. For the test, 50 µl of ¹²⁵I-labelled IPLSm1 IgG2a was incubated with 50 µl of sera at a dilution of 1/25 in PBS-0.1% BSA buffer for 1 h at 37 °C and for 16 h at 4 °C and the plates were then washed three times in PBS-0.1% BSA buffer. The wells were counted in a γ-counter. The percentage of IPLSm1 inhibition binding was calculated using the following formula: $a - b/a \times 100 = \% \text{ of inhibition}$, where $a = \text{c.p.m. obtained when } ^{125}\text{I-AB}_1 \text{ was incubated with 50 } \mu\text{l of PBS-0.1 \% BSA buffer}$, and $b, \text{ c.p.m. obtained when } ^{125}\text{I-AB}_1 \text{ was incubated with 50 } \mu\text{l serum from a rat bearing a subcutaneous IPLSm1 hybridoma (□), 50 } \mu\text{l serum from a rat immunized with AB}_2 \text{ (JM8-36) (●), 50 } \mu\text{l serum from a rat immunized with normal IgM (□), 50 } \mu\text{l serum from a 4-week-infected rat (■) or 50 } \mu\text{l of normal rat serum (■) (mean of five duplicate experiments } \pm \text{ s.d.)}$.

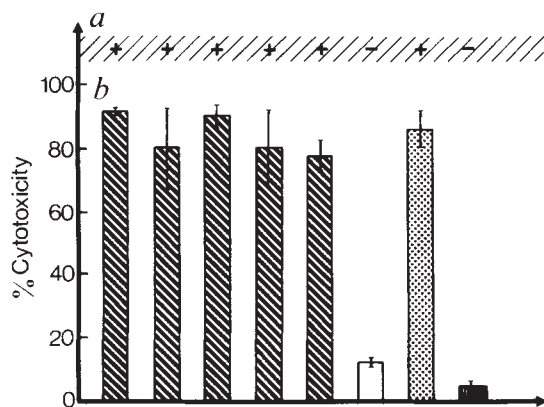


Fig. 2 Humoral response to AB₂ immunization. *a*, Immunofluorescence tests; *b*, eosinophil-dependent cytotoxicity. Indirect immunofluorescence tests were performed¹ on deep-frozen sections of schistosomula (8 μ m). *S. mansoni* schistosomula were prepared by penetration of cercariae through abdominal mouse skin. Sera (tested at a dilution of 1/20 in 10 mM PBS) were collected from LOU/M rats immunized by JM8-36 (AB₂) antibody (▨) purified from ascitic fluid as described previously¹⁵. A 1-mg aliquot of the purified preparation in physiological saline was injected subcutaneously at 2-week intervals. Control sera were obtained from rats injected in the same conditions with the IgM fraction of normal rat serum (□). Eosinophil-dependent cytotoxicity was measured after 48 h contact of skin schistosomula, preincubated overnight with the sera at a final dilution of 1/16, with a rat eosinophil-rich population (40–60% eosinophils). The percentage cytotoxicity was compared at equivalent dilutions with control sera, normal rat serum (▨) or serum from rats infected for 4 weeks with *S. mansoni* (▩). All results correspond to a mean of two duplicate experiments $\pm$ s.d.

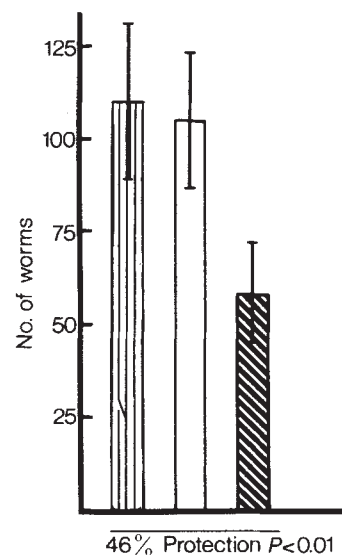


Fig. 3 Passive transfer of AB₃ antibodies. Serum (1 ml) collected from JM8-36 immunized rats (▨) or from rats immunized with normal IgM (□) was injected intravenously into LOU rats infected 4 h previously with 800 *S. mansoni* cercariae. Parasitic burdens were evaluated 3 weeks later by liver perfusion¹⁶. The number of worms obtained from these groups was compared with the parasite burden of rats injected with 1 ml of physiological saline (▨). Percentage protection was calculated by the formula $(a - b)/a \times 100$, where *a* = the number of worms recovered from the saline-injected control group and *b* = the number of worms recovered from rat injected with 1 ml of AB₃ serum.

IPLSm1 and 1 mg of aggregated human IgG in incomplete Freund's adjuvant at 2-week intervals. The Ouchterlony test was used to select rats showing the highest responses and spleen cells from these animals were fused with cells from the IR983F rat myeloma cell line⁶ as described previously¹. The hybrid cell supernatants were screened for their ability to inhibit the binding of ¹²⁵I-labelled IPLSm1 to the 38K target antigen coated on a polyvinyl chloride (PVC) plate. From 200 hybrid cell supernatants obtained from two successive cell fusion experiments, we selected 29 supernatants inducing significant levels of inhibition (>70%).

Analysis of the immunological components of these hybrid cell supernatants by the Ouchterlony test revealed that in most cases (23/29) they contained antibodies of the IgM class. The inhibitory activity of hybrid cell supernatants on IPLSm1 binding to 38K strongly suggested that the antibodies produced were able to fix to an epitope of IPLSm1 which was close to or part of its antigen-combining site, indicating that AB₂ were paratope-induced antibodies bearing an internal image of the original epitope.

We next investigated the potential use of such antibodies to induce active immunization in naive LOU rats. Two questions in particular were studied: Is it possible to use AB₂ antibodies to induce the production of AB₃ antibodies which will reproduce the effector mechanism described previously *in vitro* both in experimental schistosomiasis⁷ and with AB₁ antibody, involving eosinophil-mediated cytotoxicity¹? Does this immunization lead to a significant protection of naive rats against a challenge infection?

To answer the first question, we followed the kinetics of the antibody response in naive LOU rats immunized with AB₂ antibodies. Successive injections of purified IgM anti-idiotypic antibody (JM8-36) for 4 weeks elicited anti-schistosome antibodies (AB₃) which were detected by indirect immunofluorescence on cryopreserved *S. mansoni* schistosomulum sections. In all cases, the fluorescence reaction was observed at the surface

level. These observations, suggesting a close correlation between the structure of the anti-idiotypic antibody and an epitope present on the surface of *S. mansoni* schistosomula, were confirmed by competition of AB₁ by AB₃ on radiolabelled membrane preparations of schistosomula. AB₃ preparations strongly inhibited the binding of AB₁ to the 38K molecule, indicating a close specificity for the same epitope (Fig. 1).

We explored the biological activity of these AB₃ antibodies induced by anti-idiotypic immunization by determining their possible participation in eosinophil-dependent cytotoxicity reactions against *S. mansoni* schistosomula. A significant level of cytotoxicity (70–90%) was observed, close to that mediated by a 4-week *S. mansoni*-infected rat serum (Fig. 2) or by the AB₁ monoclonal antibody itself¹.

The relevance of these *in vitro* findings was fully supported by the demonstration that passive transfer of AB₃ antibodies to naive rats resulted in a significant protection of these animals to a challenge infection by *S. mansoni* cercariae (Fig. 3).

Finally, we studied the direct protective effect of an anti-idiotypic immunization. In two series of experiments involving 30 naive LOU rats, immunization with AB₂ resulted in 50–76% protection to a challenge infection (Fig. 4).

The work presented here clearly demonstrates that immunization with an anti-idiotypic monoclonal antibody can reproduce several parameters of the acquired immunity observed in experimental rat schistosomiasis. First, immunization with AB₂ resulted in the production of specific anti-*S. mansoni* antibodies in animals which had never been exposed to the parasite antigen. The antibodies produced exhibited the same specificity for the 38K molecule as the original AB₁ monoclonal antibody used to produce the anti-idiotypic antibody, suggesting that the AB₂ produced may represent an internal image of the original epitope. However, most importantly, such experiments raise the possibility of inducing a strong immunity against schistosomes, as shown by both active immunization and passive transfer experiments.

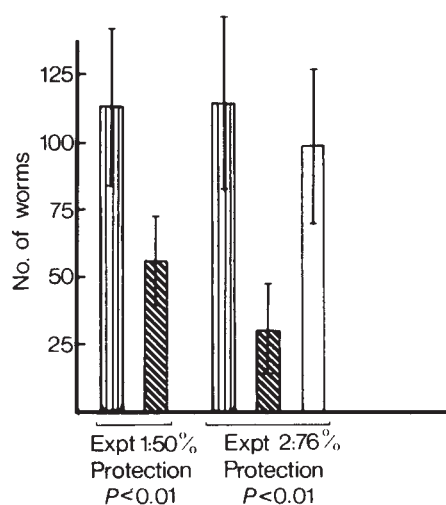


Fig. 4 Protective effect of JM8-36 immunization. LOU rats immunized as described in Fig. 2 legend, giving positive immunofluorescence reactions and significant eosinophil-dependent cytotoxicity, were infected with 800 *S. mansoni* cercariae. Parasite burdens were measured 3 weeks later by liver perfusion¹⁶. The number of worms obtained from rats immunized with JM8-36 AB₂ antibodies (■) was compared with those obtained from control groups, that is LOU rats injected with physiological saline (□) or with IgM purified from normal rat serum (▨). The percentage of protection was calculated by the formula $(a - b)/a \times 100$, where a = the number of worms obtained from the saline-injected control group and b = the number of worms recovered from AB₂-immunized rats.

Thus, immunization with anti-idiotypic antibodies represents an alternative approach to immunization against pathogens. Although recent studies have produced encouraging results⁸⁻¹² concerning the potential substitution of conventional vaccines by anti-idiotypic antibodies, this strategy is only in its early stages¹³. In the context of schistosomiasis, idiotypic vaccines could be of particular use when relevant protective epitopes cannot easily be produced by the modern tools of molecular biology. Although the rat is a non-permissive host, there is now clear evidence¹⁴ that all the specific effector mechanisms of immunity described in this model also occur in human infection. The possibility of using such immunization procedures in humans remains unexplored and cannot be directly extrapolated from experimental infections. However, work in progress in our laboratory, indicating the existence of cross-reacting idiotypes in human schistosome infection, is encouraging.

We thank Claudine Colson for her help in the preparation of the manuscript. This work was supported by INSERM U167-CNRS 624 and the WHO-UNDP Special Program for Research and Training in Tropical Diseases.

Received 18 February; accepted 24 April 1985.

- Grzych, J. M., Capron, M., Bazin, H. & Capron, A. *J. Immun.* **129**, 2739-2743 (1982).
- Dissous, C., Grzych, J. M. & Capron, A. *J. Immun.* **129**, 2232-2234 (1982).
- Dissous, C., Prata, A. & Capron, A. *J. infect. Dis.* **149**, 227-233 (1984).
- Grzych, J. M., Capron, M., Dissous, C. & Capron, A. *J. Immun.* **133**, 998-1004 (1984).
- Jerne, N. K. *Ann. Immun., Inst. Pasteur, Paris* **125C**, 373-389 (1974).
- Bazin, H., Grzych, J. M., Verwaerde, C. & Capron, A. *Ann. Immun., Inst. Pasteur, Paris* **131D**, 359 (1980).
- Capron, M. *et al. Eur. J. Immun.* **8**, 127-133 (1978).
- Sacks, D. L., Esser, K. M. & Sher, A. *J. exp. Med.* **155**, 1108-1119 (1982).
- McNamara, M. K., Ward, R. E. & Kohler, H. *Science* **226**, 1325-1326 (1984).
- Stein, K. E. & Söderström, T. *J. exp. Med.* **160**, 1001-1011 (1984).
- Kennedy, R. C., Melnick, J. L. & Dreesman, G. R. *Science* **223**, 930-931 (1984).
- Reagan, K. J., Wunner, W. H., Wiktor, T. J. & Koprowski, H. *J. Virol.* **48**, 660-666 (1983).
- Sacks, D. L. in *New Approaches to Vaccine Development* (eds Bell, R. & Torrigiani, G.) 121-131 (Schwabe, Basel, 1984).
- Capron, A., Dessaint, J. P., Capron, M. & Joseph, M. in *Progress in Immunology* Vol. 5 (eds Yamamura, Y. & Tada, T.) 1305-1316 (Academic, Tokyo, 1984).
- Bazin, H., Beckers, A. & Ouerinjeau, P. *Eur. J. Immun.* **4**, 44-48 (1974).
- Smithers, S. R. & Terry, R. J. *Parasitology* **55**, 695-700 (1965).

Individual-specific 'fingerprints' of human DNA

A. J. Jeffreys*, V. Wilson* & S. L. Thein†

* Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

† MRC Molecular Haematology Unit, Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

Simple tandem-repetitive regions of DNA (or 'minisatellites') which are dispersed in the human genome frequently show substantial length polymorphism arising from unequal exchanges which alter the number of short tandem repeats in a minisatellite¹⁻⁴. We have shown previously that the repeat elements in a subset of human minisatellites share a common 10-15-base-pair (bp) 'core' sequence which might act as a recombination signal in the generation of these hypervariable regions⁵. A hybridization probe consisting of the core repeated in tandem can detect many highly polymorphic minisatellites simultaneously to provide a set of genetic markers of general use in human linkage analysis⁵. We now show that other variant (core)_n probes can detect additional sets of hypervariable minisatellites to produce somatically stable DNA 'fingerprints' which are completely specific to an individual (or to his or her identical twin) and can be applied directly to problems of human identification, including parenthood testing.

Three human minisatellites, termed 33.5, 33.6 and 33.15, each comprised of tandem repeats of various versions of the core sequence, have been cloned previously and characterized⁵ (Fig.

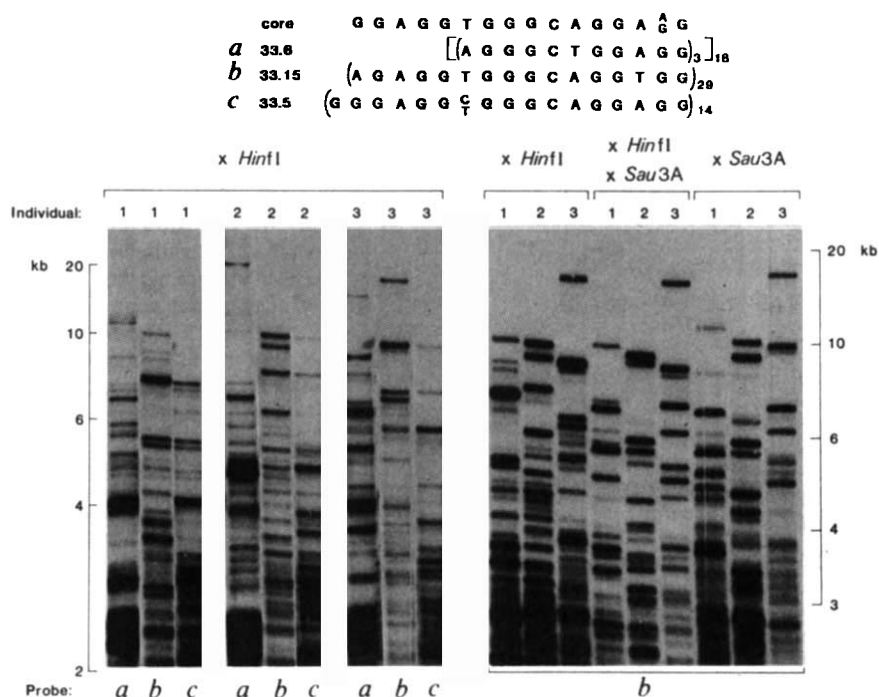
Table 1 Similarities of DNA fingerprints between random pairs of individuals

Probe	DNA fragment size (kb)	No. of fragments per individual $\pm$ s.d.	Probability x that fragment in A is present in B	Maximum mean allelic frequency/homozygosity
33.6	10-20	2.8 ± 1.0	0.11	0.06
	6-10	5.1 ± 1.3	0.18	0.09
	4-6	5.9 ± 1.6	0.28	0.14
33.15	10-20	2.9 ± 1.0	0.08	0.04
	6-10	5.1 ± 1.1	0.20	0.10
	4-6	6.7 ± 1.2	0.27	0.14

Samples (8 μ g) of blood DNA⁶ taken from a random sample of 20 unrelated British caucasians were digested with *HinfI* and Southern blot hybridized with minisatellite probes 33.6 or 33.15 as described in Fig. 1 legend. Each DNA fingerprint (individual A) was compared with the pattern in the adjacent gel track (individual B), and the number of bands in A which were clearly absent from B, plus those which had a co-migrating counterpart of roughly similar autoradiographic intensity in B, were scored. The data shown are averages for all pairwise comparisons. A small proportion (~6%) of additional weakly hybridizing fragments in A were matched by strongly hybridizing fragments in B, and because in such cases it was not possible to decide whether the band in A was also present in B, such fragments were ignored. If co-migrating bands in A and B are always identical alleles of the same minisatellite locus, then the probability x that an allele in A is also present in B is related to the frequency q of that allele by $x = 2q - q^2$. As the allele frequency is low, then $q^2 \ll q$ and therefore the mean probability $\hat{x}$ is approximately related to the mean allele frequency $\hat{q}$ by $\hat{x} \approx 2\hat{q}$. Furthermore, assuming that there is little variance in q between alleles, then the number of alleles $n = 1/\hat{q}$ and the mean homozygosity is therefore approximately given by $\sum_1^n q_i^2 = n\hat{q}^2 = \hat{q}$. In practice, an (unknown) proportion of co-migrating bands in A and B will be derived by chance from different minisatellite loci, and thus the estimates of mean allele frequency and homozygosity are maximal and depend on the electrophoretic resolution of minisatellite fragments. Probability estimates: the mean probability that all fragments detected by probe 33.15 in individual A are also present in B is $0.08^{2.9} \times 0.20^{5.1} \times 0.27^{6.7} = 3 \times 10^{-11}$.

Fig. 1 Hypervariable fingerprints of human DNA. DNA samples prepared from three individual placentae (1-3) were digested with *Hinf*I and/or *Sau*3A and Southern blot hybridized with ³²P-labelled single-stranded DNA probes prepared from M13 recombinants 33.5, 33.6 or 33.15, each of which contains a human minisatellite consisting of tandem repeats of closely related variants of the consensus sequences shown; the repeat unit in 33.6 is in turn a diverged trimer⁵. Each probe produces a different fragment pattern whose complexity is largely independent of the tetranucleotide restriction endonuclease used. Resolution of polymorphic fragments less than 4 kb long is improved in double digests with *Hinf*I plus *Sau*3A, due to the elimination of background hybridization caused presumably by relatively diverged and invariant *Hinf*I minisatellite fragments⁵ which have accumulated *Sau*3A cleavage sites within one or more repeat units. In double digests, the number of resolvable polymorphic fragments detected by probe 33.15 can be increased from ~15 to ~23 per individual, at the expense of losing ~20% of long single-digest minisatellite fragments which presumably contain a *Sau*3A cleavage site in most or all repeat units.

Methods. DNA was isolated from fresh human placentae as described elsewhere⁶. Samples (8 µg) of DNA were digested with *Hinf*I and/or *Sau*3A, in the presence of 4 mM spermidine trichloride to aid complete digestion, recovered after phenol extraction by ethanol precipitation, and electrophoresed through a 20-cm long 0.6% agarose gel at 30 V for ~24 h, until all DNA fragments <1.5 kb long had electrophoresed off the gel. DNA was then transferred by blotting to a Sartorius nitrocellulose filter⁷. High specific activity (>10⁹ c.p.m. ³²P per µg DNA) single-stranded M13 probes were prepared as described previously⁵. The precise probes used were: 33.5, a 220-nucleotide *Hae*III fragment containing the minisatellite plus 60 nucleotides of flanking human DNA, subcloned into the *Sma*I site of M13mp8 (ref. 8); 33.6, a 720-nucleotide *Hae*III fragment containing the minisatellite plus 50 nucleotides of flanking human DNA subcloned into the *Sma*I site of M13mp8; 33.15, a 592-nucleotide *Pst*I/*Aha*III fragment containing the minisatellite plus 128 nucleotides of flanking human DNA subcloned into M13mp19 DNA digested with *Pst*I plus *Sma*I. Southern blot hybridization and washing were performed in 1 × SSC at 65 °C as described previously⁵. Filters were autoradiographed at room temperature without an intensifier screen for 4 days.



1). Probe 33.15 has been shown to hybridize to multiple hypervariable fragments in *Hinf*I digests of human DNA⁵. Probes 33.5 and 33.6 also detect a complex set of hypervariable regions in human DNA (Fig. 1). Probes 33.5 and 33.15 contain repeats of a similar version of the complete core sequence and consequently produce similar, but not identical, DNA fingerprints. In contrast, probe 33.6 is comprised of a shortened derivative of the core and hybridizes to a largely novel set of hypervariable fragments. Probe 33.15 detects ~15 resolvable hypervariable fragments per individual in the 4–20-kilobase size range, whereas probe 33.6 detects ~11 additional fragments of this size which are not detected by probe 33.15. Probe 33.5 hybridizes on average to about two further large fragments not detected by probes 33.6 or 33.15.

The DNA fingerprint pattern for the longest hypervariable fragments is largely independent of the 4-bp recognition restriction endonuclease used (Fig. 1). This strongly suggests that these large fragments are not derived from longer minisatellites, but that each contains a complete long homogeneous minisatellite devoid of restriction endonuclease cleavage sites and flanked by human DNA containing the normal high density of 4-bp cleavage sites. This is in agreement with previous data showing that most of these large minisatellite fragments are unlinked and segregate independently in pedigrees⁵.

To determine the variability of DNA fingerprints, DNA samples from a panel of 20 unrelated British caucasians were digested with *Hinf*I and fingerprinted using probes 33.6 and 33.15 (Table 1). Pairwise comparisons of the DNA fingerprints showed that the minisatellite patterns were highly specific to an individual, and that few fragments are shared between two randomly selected individuals. The probability of shared bands increases for smaller minisatellite fragments, probably resulting from lower genetic variability (higher allele frequencies) of these loci⁵ combined with the fortuitous co-migration of unrelated

minisatellite fragments. From the degree of band sharing, one can obtain maximal approximate estimates of mean allele frequency and homozygosity (Table 1). For the longest minisatellite fragments in particular, the mean allele frequency is very low (<0.04) and the mean heterozygosity rises to >96%. This is consistent with previous pedigree analysis which has shown that most of these large hypervariable DNA fragments are present in the heterozygous state⁵.

The data in Table 1 allow us to estimate the individual specificity of a DNA fingerprint. For probe 33.15, the probability that all 15 resolved fragments in the 4–20-kb size range in an average individual A are also present in a second unrelated individual B is 3×10^{-11} (see Table 1 for details of probability estimates); the probability that the fingerprints of A and B are identical, that is, that all fragments less than 4 kb also match and that B does not possess any additional 4–20-kb fragments, is therefore $\ll 3 \times 10^{-11}$. Similarly, the probability that A and B have identical fingerprints for both probes 33.15 and 33.6 is $\ll 5 \times 10^{-19}$. If individuals A and B are related, the chance of fingerprint identity is increased, to a maximum for parents/offspring and sibs. For nonconsanguineous marriages, as the heterozygosity for each fragment is very high, the probability that a hypervariable fragment in sib A is present in sib B is ~1/2, and thus the probability that all 15 resolved bands detected by probe 33.15 in sib A are present in sib B is $\sim 2^{-15} = 3 \times 10^{-5}$ ($\sim 10^{-8}$ for both probes 33.15 and 33.6). These DNA fingerprints are therefore almost totally individual-specific, even within a single family (except for identical twins, see below).

The DNA fingerprints obtained using these core minisatellite probes are reproducible and are suitable for individual identification. In addition, sufficient DNA (0.5–5 µg) can be isolated rapidly from a single drop of human blood for DNA fingerprinting. This is illustrated in Fig. 2, in which DNA fingerprints were produced from a randomized panel of individuals, including

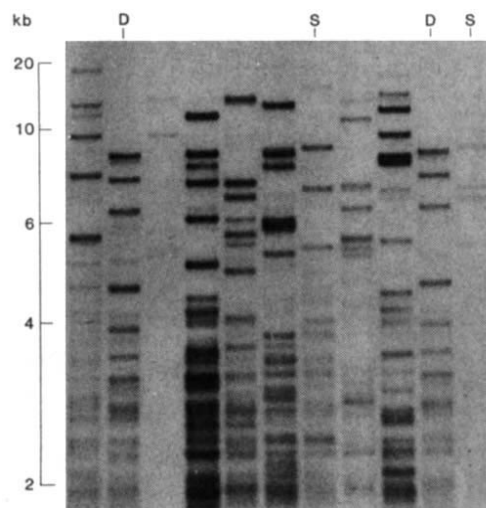


Fig. 2 Individual identification using DNA fingerprints from small samples of blood. DNA was prepared by a rapid procedure using one or two drops of blood from a panel of 11 individuals and digested with *Hinf*I; DNA fingerprints were prepared as described in Fig. 1 legend, using 33.15 as a probe. The panel consisted of nine unrelated individuals, two of whom had been previously DNA fingerprinted, plus two sisters. The autoradiograph was inspected by a colleague who was unaware of the order of samples on the autoradiograph. He correctly identified the two previously fingerprinted individuals on the basis of pattern identity, as well as identifying the two sisters, who have several bands in common (tracks S). He also correctly deduced the fact that duplicate samples had been taken from another individual (tracks D). **Methods.** One or two drops (30–100 μ l) of blood were collected from a fingerprick into a 1 ml 1 $\times$ SSC in an Eppendorf centrifuge tube. Cells were pelleted in an Eppendorf microfuge for 1 min, haemolysed by rapid suspension in 1 ml water and immediately made isotonic by the addition of 0.25 ml 5 $\times$ SSC. White blood cells, nuclei and red cell ghosts were pelleted by centrifugation for 3 min, resuspended in 0.2 ml 0.2 M Na-acetate pH 5.6 and lysed by the addition of SDS to 1%. The lysate was extracted twice with phenol/chloroform and DNA collected by two rounds of ethanol precipitation at room temperature. DNA was dissolved in 20 μ l water and digested with *Hinf*I in the presence of 4 mM spermidine trichloride at 37°C for 1 h. Electrophoresis and hybridization were performed as described for Fig. 1.

two people whose DNA had been fingerprinted previously, and two sisters. The two previously characterized individuals could be readily and unambiguously identified on the basis of DNA fingerprint comparisons, as could the two sisters, who have a substantial number of minisatellite fragments in common.

We have shown previously that these hypervariable minisatellite fragments are stably inherited in a mendelian fashion, and that the mutation rate to a new length allele is low (of the order of 0.001–0.004 per locus per gamete for the longest minisatellite fragments)⁵. Several experiments show that these DNA fingerprints are also somatically stable, an essential prerequisite for identification purposes (Fig. 3). Thus, the DNA fingerprints for sperm and blood DNA are indistinguishable, as are the patterns of monozygous twins. Furthermore, the patterns appear to be stably maintained in cultured cells, as shown by comparing the DNA fingerprints of blood DNA with DNA isolated from Epstein-Barr virus-transformed lymphoblastoid cell lines derived from the same individual.

The DNA fingerprints produced by minisatellite probes 33.6 and 33.15 are therefore sufficiently stable and individual-specific for use in human identification in, for example, forensic medicine, and could be used for the routine identification and authentication of human cell lines in culture. They also provide a reliable method for paternity testing (see Fig. 3). Approximately half of the polymorphic minisatellite fragments in an offspring are derived from the father, and these paternal frag-

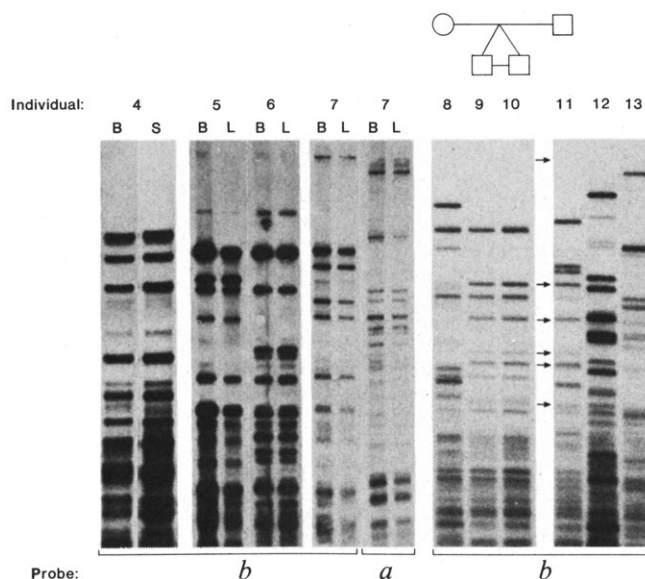


Fig. 3 Somatic stability of DNA fingerprints and their use in paternity testing. Patterns of hypervariable DNA fragments are compared between blood (B) and sperm DNA (S) of individual 4, and between blood DNA (B) and DNA isolated from transformed lymphoblastoid cell lines (L) derived from related individuals 5–7 (6 is the daughter and 7 is the maternal aunt of woman 5; the kinship is evident in the number of fragments shared by these individuals). DNA fingerprints are also shown for blood DNA from identical twins (9, 10) and are compared with their mother (8), father (11) and two unrelated men (12, 13). Resolved paternal DNA fragments in the twins (arrowed) were identified by eliminating maternal bands, and are all present in the father but not in individuals 12 or 13.

Methods. Lymphoblastoid cell lines transformed by Epstein-Barr virus⁹ and stored in liquid nitrogen were re-established in liquid culture after 2 yr. These cultured lymphocytes were washed twice in normal saline; DNA from the lymphocyte pellet and from white blood cells was prepared as described elsewhere⁶. Sperm DNA was prepared similarly, except that sperm collected from semen were treated with 1 M 2-mercaptoethanol for 5 min at room temperature before lysis with SDS. DNA fingerprints were prepared as described for Fig. 1, using 5- μ g samples of DNA digested with *Hinf*I and hybridized with probe 33.6 (a) or 33.15 (b).

ments can be identified by comparison of the mother's and offspring's DNA fingerprints. All of these paternal fragments must be present in the father (allowing for a possible rare new mutation). Because, in practice, probe 33.15 will detect approximately six resolved paternal fragments in the 4–20-kb range (Table 1, Fig. 3), we can use the data in Table 1 to estimate the probability that incorrect paternity will not be detected, that is, that the putative father will by chance possess all six paternal-specific DNA fragments. This probability is low ($\sim 5 \times 10^{-5}$) if the putative father is not related to the true father, but increases to $\sim 2^{-6} = \sim 0.016$ if they are closely related (brothers, father/son). If both probes 33.6 and 33.15 are used, these probabilities are reduced to $\sim 4 \times 10^{-8}$ and $\sim 8 \times 10^{-4}$, respectively, although the precise probabilities will depend on the exact resolution and complexity of the DNA fingerprint patterns obtained, and will be improved if additional paternal fragments <4 kb long can be identified (for example, by using double digests, Fig. 1). We conclude that in the vast majority of cases, the combined use of probes 33.6 and 33.15 will be sufficient to identify cases of incorrect parenthood. An interesting corollary is that these DNA fingerprints could be used with an equal level of reliability to establish true biological parentage.

This work is the subject of a UK patent application. Enquiries should be addressed to the NRDC. We thank John F. Y. Brookfield for helpful discussions, Dr R. S. Pereira for help and advice with lymphoblastoid cell lines, and Dr G. Corney and many

volunteers at Leicester for providing blood samples. A.J.J. is a Lister Institute Research Fellow, and this work was supported by a grant to A.J.J. and a Training Fellowship to S.L.T. from the MRC.

Received 6 March; accepted 3 May 1985.

1. Bell, G. I., Selby, M. J. & Rutter, W. J. *Nature* **295**, 31–35 (1982).
2. Proudfoot, N. J., Gil, A. & Maniatis, T. *Cell* **31**, 553–563 (1982).
3. Goodbourn, S. E. Y., Higgs, D. R., Clegg, J. B. & Weatherall, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5022–5026 (1983).
4. Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. *Nature* **302**, 33–37 (1983).
5. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **314**, 67–73 (1985).
6. Jeffreys, A. J. *Cell* **18**, 1–10 (1979).
7. Southern, E. M. *Meth. Enzym.* **68**, 152–176 (1980).
8. Vieira, J. & Messing, J. *Gene* **19**, 259–268 (1982).
9. Nilsson, K., Klein, G., Henle, W. & Henle, G. *Int. J. Cancer* **8**, 443–450 (1971).

DNA restriction fragments associated with α_1 -antitrypsin indicate a single origin for deficiency allele *PIZ*

Diane Wilson Cox, Savio L. C. Woo* & Tammy Mansfield

Research Institute, Hospital for Sick Children, 555 University Avenue, Toronto, Canada M5G 1X8 and Departments of Paediatrics, Medical Genetics & Medical Biophysics, University of Toronto, Toronto, M5S 1A8, Canada

* Howard Hughes Medical Institute, Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030, USA

The α_1 -protease inhibitor, or α -antitrypsin (AAT), a major plasma inhibitor of leukocyte elastase and bacterial proteases, is encoded at the *PI* locus on chromosome 14 (14q24.3–q32.1)¹. A deficiency of AAT in individuals homozygous for the *PIZ* allele occurs in about 1 in 2,000–8,000 caucasians² and is associated with an increased risk of early adult onset emphysema³ and liver disease in childhood⁴. We have now used DNA polymorphisms associated with the AAT gene to investigate the origin of the *PIZ* allele. Using two genomic probes⁵ extending into the 5' and 3' flanking regions, respectively, we have identified eight polymorphic restriction sites. Extensive linkage disequilibrium occurs throughout the probed region with the *PIZ* allele, but not with normal *PIM* alleles. The *Z* allele occurs mainly with one haplotype, indicating a single, relatively recent, origin in caucasians.

Venous blood samples were collected in EDTA from 32 normal unrelated controls, 26 patients with AAT deficiency (PI type ZZ), and 15 pairs of parents and 16 sibs of patients with AAT deficiency. PI type ZZ patients included 12 adults with emphysema, 3 adults with liver disease, 9 children with liver disease and 2 healthy adults. Relatives of 16 normal controls were studied: members of 2 normal two-generation families with a total of 7 children, and 53 members of a three-generation family of a PI ZZ proband which included 12 unrelated spouses. The genetic type of AAT (PI type) was determined by isoelectric focusing of plasma in acrylamide gels, pH 3.5–6 (ref. 6). Blot hybridization of leukocyte DNA digested with specific enzymes

is described in Fig. 2 legend. The 5' probe (4.6 kilobases, kb) and 3' probe (6.5 kb) have been described elsewhere⁷ (Fig. 1), both having been subcloned from phage clone α AT35⁵.

Of 10 restriction enzymes tested with the 5' 4.6-kb probe, 3 showed restriction site polymorphisms; these were *Sst*I, *Msp*I and *Ava*II. For each enzyme, one or both of two possible restriction fragments were observed. When a particular restriction site was present (designated +), a shorter DNA fragment was observed than when the restriction site was absent (designated –). Family studies indicated that the pair of DNA fragments for each enzyme were alleles. Among 32 unrelated normal controls, the allele frequencies were as shown in Table 1. The 4.6-kb genomic probe has been completely sequenced⁸, so the positions of the restriction sites were identified. Fragments of the expected sizes were observed. The distances between restriction sites were 2 kb between *Sst*I and *Msp*I sites, and 0.2 kb between *Msp*I and *Ava*II sites.

A χ^2 analysis of 2×2 tables, or exact calculation of probabilities where required⁹, was used to examine the data for possible association between a specific allele at one site and a specific allele at each of the other two sites. No significant association between alleles was observed, indicating free recombination between the three restriction sites, with no evidence



Fig. 1 The α_1 -antitrypsin gene and adjacent flanking regions, described previously^{4,17}. Solid areas are coding regions, dots indicate untranslated regions, open areas are introns. Restriction sites for the enzymes *Eco*RI and *Bam*HI (B) are indicated. *, Polymorphic restriction sites for *Sst*I (S), *Msp*I (M) and *Ava*II (A). The two genomic probes used in this study are shown above the gene.

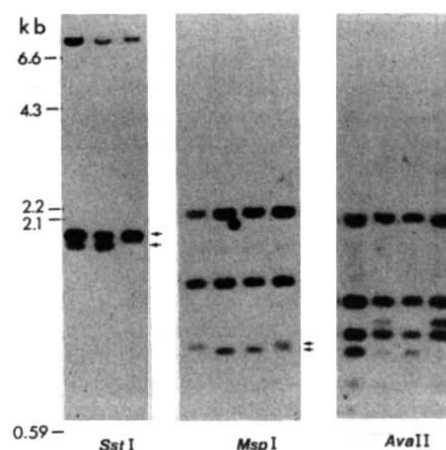


Fig. 2 Patterns of DNA fragments are shown for the three enzymes *Sst*I, *Msp*I and *Ava*II, using the 4.6-kb 5' probe. DNA extracted from leukocytes was completely digested with specific restriction enzymes according to conditions recommended by the manufacturer. Fragments were separated by size in 0.8% agarose gels. DNA fragments were transferred to Biotrans membranes (Pall). Probes were labelled with ³²P using a nick-translation kit (Amersham). After prehybridization, ³²P-labelled probe (3.0×10⁷ counts per filter) was hybridized overnight. Filters were rinsed twice in 2×SSC (SSC: 1.5×10⁻⁴ M NaCl, 1.5×10⁻⁵ M Na-citrate). Membranes were washed in 2×SSC at 60 °C for 45 min, then twice in 0.1% SDS in 0.1×SSC at 60 °C for 45 min, rinsed briefly in SSC, blotted, then exposed to X-ray film for 2–3 days. Fragment sizes in kb are indicated on the left. Anode is at the bottom. Arrows mark the polymorphic fragments. From left to right, lanes show the following allele designations: ++, ++, --; --, ++, ++, ++; ++, ++, ++, --.

Table 1 Allele frequencies for 5' DNA polymorphisms (4.6-kb probe)

Restriction enzyme	Alleles (kb)	Allele frequency +	Allele frequency –	PIC*
<i>Sst</i> I	1.8, 1.9	0.649	0.351	0.35
<i>Msp</i> I	0.95, 0.98	0.506	0.494	0.38
<i>Ava</i> II	0.9, 1.1	0.636	0.364	0.36

* Polymorphism information content, calculated according to the formula of Botstein *et al.*²¹, expressing the extent of variability at each locus.

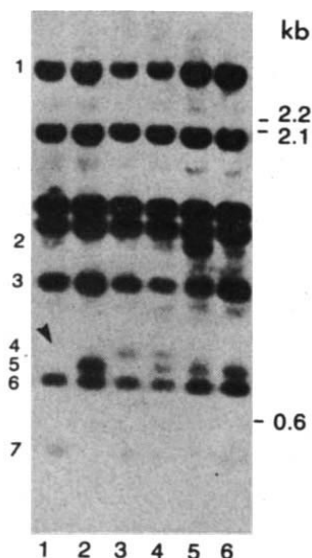


Fig. 3 DNA fragments observed using restriction enzyme *Ava*II and the 6.5-kb probe. Sizes of fragments are indicated in kb, anode is at the bottom. The seven polymorphic fragments are numbered on the left. There are also three strong bands present between polymorphic bands 1 and 2 which appear to be constant in all samples observed to date. The *Ava*II DNA phenotypes observed in each lane are as follows: (1) 11336677; (2) 11336655; (3) 10334677; (4) 10334657; (5) 11236657; (6) 11336657. The arrow in lane 1 shows the unique region of the pattern associated with PI type ZZ.

of linkage disequilibrium between them in normal individuals, despite the close proximity of sites. The specific combination of the three restriction sites (DNA haplotype) was determined by segregation within families. Forty-seven non-Z haplotypes (M of various subtypes and S) could be determined from 16 normal controls and from 15 parents of PI ZZ children using only the non-Z haplotype. The specific haplotypes found with each PI type are shown in Table 2. The expected haplotypes for all non-Z PI types were calculated on the assumption that each restriction site polymorphism segregates independently, so that the probability of a particular haplotype occurring is the product of each of the component allele frequencies. There is no significant difference between observed and expected frequencies ($\chi^2 = 10.7$; d.f. = 7) for the data available (Table 2), although the $-++$ and $-+-$ categories may be decreased in larger series. Again, the conclusion is that there has been considerable recombination between the three restriction sites, with no linkage disequilibrium for non-Z haplotypes. The DNA haplotypes found in combination with specific PI types are shown in Table 2. Further data will be required to determine whether there is linkage disequilibrium between each specific PI type and specific haplotypes, although our data suggest there is disequilibrium for PI variants M3 and S.

In contrast to the results for the normal non-Z PI types, among the 26 unrelated individuals with AAT deficiency (PI type ZZ), 25 were homozygous for the haplotype *Sst*I $^-$, *Msp*I $^+$, *Ava*II $^+$. The haplotype $-++$ therefore occurred with 51 of 52 PI Z alleles. One PI Z allele was associated with the 5' DNA haplotype $+++$, indicating a crossover event between the *Sst*I and *Msp*I sites, or a mutation at the *Sst*I site. There is therefore a very high degree of linkage disequilibrium for the 5' restriction sites of PI Z alleles. This specific haplotype ($-++$) is found in association with 8 of 28 PI M1 alleles and 1 of 13 M2 alleles, suggesting that the PI Z allele has been derived from the PI M1 allele. The amino-acid sequence of the Z AAT product differs in one position from that of M, subtype undefined: at position 342 in exon V, glutamine in M is changed to lysine in Z^{10,11}.

Further polymorphisms were found using the 3' 6.5-kb probe and the restriction enzyme *Ava*II. The pattern of DNA fragments

included at least seven polymorphic fragments and three constant fragments, ranging in size from 0.48 to 2.7 kb (Fig. 3). Because sequencing is incomplete for the region of the 6.5-kb probe, not all specific sites have been identified, so fragments have been numbered. The polymorphic band 1 (2.7 kb) was either present or absent, indicated as alternate alleles 0 or 1, with frequencies of 0.36 and 0.64, respectively. The fragment labelled 2 (1.15 kb) (Fig. 3, lane 5) represents a rare allele, frequency 0.02. Bands 4 (0.72 kb) and 6 (0.64 kb) are alleles, frequencies 0.30 and 0.70, respectively. At least one constant fragment is also present in position 6. Alleles represented by bands 5 (0.68 kb) and 7 (0.48 kb) have frequencies of 0.07 and 0.30, respectively. Fragment 7 transfers inefficiently and can usually be seen clearly only when present in the homozygous state. The proposed interpretation of alleles is supported by family studies.

The phenotype observed in lane 1 of Fig. 3, with both bands 4 and 5 absent, was not observed in 32 normal control individuals. However, this band pattern was observed in all of 26 individuals of PI type ZZ. The particular combination of *Ava*II DNA fragments can be stated as a haplotype. PI type ZZ individuals are therefore homozygous for the haplotype 1367. Of the 47 non-Z haplotypes determined by family studies, none has the 1367 haplotype. The most common *Ava*II haplotype is 1365, found with 27 of 47 non-Z alleles (11/28 M1, 10/13 M2, 4/4 M3 and 2/2 S). This haplotype would require only one mutation to produce an *Ava*II restriction site to convert fragment 5 to fragment 7. About 75% of the region of the probe has been sequenced⁸ and fragments of the sizes of polymorphic bands 2, 3, 4, 5, 6 and 7 are located within the AAT gene. Two additional restriction site polymorphisms, one within the gene and one in the 3' flanking region, were observed using the 3' probe and the restriction enzyme *Taq*I (data not shown). The polymorphic *Taq*I sites are identical for 51 of 52 Z haplotypes; the exception involves a *Taq*I site 3' to the AAT gene.

The finding of a single unique DNA haplotype for most (96%) PI Z alleles is exceptional for diseases studied to date. The first association of a DNA restriction fragment polymorphism with a disease was described for sickle-cell anaemia¹². Multiple restriction sites associated with the sickle-cell haemoglobin gene indicate three independent origins in Africa¹³. For thalassaemia, a specific restriction fragment haplotype is generally associated with a specific mutant gene and haplotypes tend to be characteristic for defined ethnic groups¹⁴. In contrast, for AAT, the mutant gene is almost always associated with a single haplotype, indicating a single origin for the PI Z allele. While most patients in the present study are of British origin, others are Ukrainian, German, Dutch and French. Many different population groups have now been surveyed², and have shown that the PI Z allele is limited to caucasians and is not found in oriental or black population groups, except where it can be explained by racial admixture. Also, there is a suggestion of a decreasing gene frequency from north to south in European populations. The

Table 2 DNA restriction fragment haplotypes (5' 4.6-kb probe) associated with specific PI types

Haplotype			Total non-Z		PI type*				
<i>Sst</i> I	<i>Msp</i> I	<i>Ava</i> II	Observed	Expected	M1	M2	M3	S	Z
-	-	-	6	3.0		6			
-	-	+	1	5.2	1				
-	+	+	9	5.3	8	1			51
-	+	-	1	3.0	1				
+	-	-	5	5.5	1	4			
+	-	+	11	9.6	9			2	
+	+	-	5	5.6	5				
+	+	+	9	9.8	3	2	4		1
Totals			47	47	28	13	4	2	52

* Controls were selected before PI typing; frequencies of PI types therefore reflect population frequencies²².

PIZ allele has certainly arisen since the divergence of the races, and probably much more recently. The number of generations which have elapsed since the origin of the *PIZ* mutation can be estimated, assuming that meiotic crossing-over is a random event (which is not strictly true) and that on average there is roughly one meiotic crossover per 10^8 base pairs per generation¹⁵. As the crossover frequency is 1 in 52, and the estimated distance between the most 5' restriction site (*SstI*) and the most 3' *AvaII* polymorphic site is approximately 9×10^3 base pairs, it can be calculated that 216 generations, or 6,480 yr, would be required to generate 1.9% crossovers. Because no crossovers have occurred in the 7-kb distance between the 5' *MspI* site and the most 3' *AvaII* site, the number of generations may be considerably less. The lack of linkage disequilibrium between the three 5' polymorphic restriction sites suggests an ancient origin for the polymorphisms, or a region of high recombination. Differential rates of recombination have been shown within the β -globin¹⁶ and immunoglobulin¹⁷ clusters. After the single

origin in a northern caucasian race, selective forces may have increased the frequency of the *PIZ* allele (see ref. 2 for a review). Such selective forces may be direct; AAT may be directly involved in the fertilization process and the decreased concentration in *PIZ* heterozygotes could enhance fertility^{18,19}.

In addition to providing information on the origin of the *PIZ* allele, the studies reported here show that various clinical manifestations of the disease are associated with only one mutant allele. Furthermore, the unique *AvaII* haplotype provides a basis for prenatal diagnosis²⁰.

We thank Drs Vincent Kidd and Torben Bech-Hansen, respectively, for subcloning the probes and assisting in probe preparation; Diane Wills for PI typing and DNA extraction; Patricia Zavitz for obtaining blood samples; and Beverly Bessey for typing the manuscript. Many physicians have provided access to their patients and their contribution is appreciated. This study was supported in part by grants from the MRC of Canada (MA5426) to D.W.C. and from NIH (HL27509) to S.L.C.W.

Received 23 January; accepted 7 May 1985.

1. Cox, D. W., Markovic, V. D. & Teshima, I. E. *Nature* **297**, 428-430 (1982).
2. Fagerhol, M. K. & Cox, D. W. *Adv. hum. Genet.* **11**, 1-62 (1981).
3. Laurell, C.-B. & Eriksson, S. *Scand. J. clin. Lab. Invest.* **15**, 132-140 (1963).
4. Sharp, H. L., Bridges, R. A., Krivit, W. & Frier, E. F. *J. Lab. clin. Med.* **73**, 934-939 (1969).
5. Leicht, M. *et al. Nature* **297**, 655-659 (1982).
6. Cox, D. W. *Am. J. hum. Genet.* **33**, 354-365 (1981).
7. Kidd, V. J., Wallace, R. B., Itakura, K. & Woo, S. L. C. *Nature* **304**, 230-234 (1983).
8. Long, G. L., Chandra, T., Kurachi, K., Woo, S. L. C. & Davie, E. W. *Biochemistry* **23**, 4828-4837 (1984).
9. Fisher, R. A. in *Statistical Methods for Research Workers*, 78-97 (Oliver and Boyd, Edinburgh, 1944).
10. Jeppson, J.-O. *FEBS Lett.* **65**, 195-197 (1976).
11. Yoshida, L., Lieberman, J., Gaidulis, L. & Ewing, C. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1324-1328 (1976).

12. Kan, Y. W. & Dozy, A. M. *Lancet* **ii**, 910-911 (1978).
13. Pagnier, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 1771-1773 (1984).
14. Orkin, S. H. *et al. Nature* **296**, 627-631 (1982).
15. Kurmit, D. M. *Lancet* **i**, 104 (1979).
16. Antonarakis, S. E., Boehm, C. D., Giardina, P. J. V. & Kazanian, H. H. Jr. *Proc. natn. Acad. Sci. U.S.A.* **79**, 137-141 (1982).
17. Bech-Hansen, N. T., Linsley, P. S. & Cox, D. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6952-6956 (1983).
18. Fagerhol, M. K. in *Proc. IVth int. Congr. hum. Genet.*, Paris (eds de Grouchy, J., Ebling, F. J. G. & Henderson, I. W.) 277-285 (Excerpta Medica, Amsterdam, 1971).
19. Kueppers, F. in *Pulmonary Emphysema and Proteolysis* (ed. Mittman, C.) 133-137 (Academic, New York, 1972).
20. Cox, D. W. & Mansfield, T. *Lancet* **i**, 230 (1985).
21. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. *Am. J. hum. Genet.* **32**, 314-331 (1980).
22. Cox, D. W. & Huber, O. *Clin. Genet.* **17**, 153-160 (1980).

Spontaneous excision of a large composite transposable element of *Drosophila melanogaster*

W. Chia*, S. McGill, R. Karp*, D. Gubb & M. Ashburner

Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK

The *TE1* family of transposable elements (TEs) of *Drosophila* consists of unusually large transposons, cytologically visible in larval polytene chromosomes as one or more bands¹⁻³. They are composite elements, as their termini consist of foldback (FB) sequences which are themselves transposable⁴⁻⁶. The location of FB elements at the termini of transposable elements suggests that these sequences have a direct role in the genetic instability of TEs. To investigate the structural and phenotypic consequence of TE excision, we have cloned genomic DNA required for the expression of the *no-ocelli* (*noc*) gene of *Drosophila*; this gene has been mutated by the insertion of *TE146*, a member of the *TE1* family carrying six polytene chromosome bands including functional copies of the *white* (*w*⁺) and *roughest* (*rst*⁺) genes. As reported here, our experiments indicate that the spontaneous excision of *TE146*, which results in the loss of the *w*⁺ and *rst*⁺ markers, can occur either as a single-step event or following a partial internal deletion. In either case, the end product is an imprecise excision in which a residual portion of the element, varying in size from 3 to 10 kilobases (kb), is left at the insertion site. These residual sequences share homology with the FB family. Furthermore, despite their imprecise nature, all these spontaneous excisions restore a wild-type *noc*⁺ phenotype.

The first transposable element identified arose from the spontaneous transposition of two adjacent genes, *w*⁺ and *rst*⁺, from

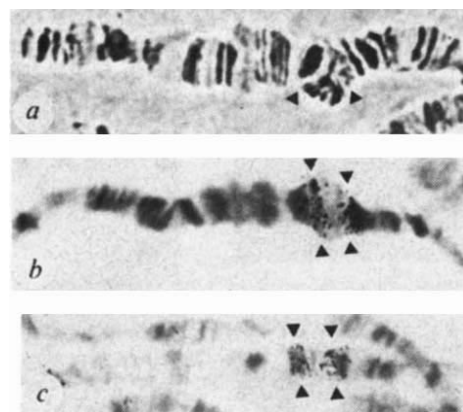


Fig. 1 *a*, Polytene chromosomes of a *TE146/+* heterozygote, showing the six bands of the TE inserted near bands 35B1.2. Arrowheads indicate the cytological limits of *TE146*. *b*, Polytene chromosomes of homozygous *TE146* hybridized with a nick-translated probe prepared from a clone (M365; ref. 6) including sequences from the *w*⁺ gene. There are two bands of homology to the probe within the TE and one at 3C1, the normal locus of *w*⁺. Arrowheads indicate sites of hybridization within *TE146*. *c*, Polytene chromosomes of homozygous *TE146* hybridized with a nick-translated probe prepared from a clone (FB8) including sequences of the repetitive FB element^{4,5}. Note that there are at least two discrete sites of hybridization within the TE (arrows).

their normal location on the X chromosome to chromosome arm 2R. The element (*TE1*) resulting from this original transposition is mobile and its derivatives have been mapped to over 150 cytological locations throughout the *Drosophila* genome².

Transposable elements can induce mutations in genes that lie at, or near, their insertion sites; for some, but not all, of these TE-induced mutations, the subsequent loss of the element leads to a restoration of the wild-type phenotype⁷. To understand this observation, and to clarify the role of FB sequences in TE instability, we examined the structural consequences of TE loss. For this purpose, *TE146*, a member of the TE family which

* Present addresses: Department of Biochemistry, University of Bristol, Bristol BS8 1TD, UK (W.C.); Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138, USA (R.K.).

Fig. 2 a, Polytene chromosomes of TE146/CyO hybridized with a genomic clone, λ ob 9.04, that crosses the site of insertion of the TE. The single site of hybridization to the wild-type homologue (open arrowhead) is split into two by the insertion of the TE (solid arrowheads). Cy indicates the left breakpoint of the *In(2L)Cy* component of *CyO*. **b**, Restriction enzyme site map of a 25-kb genomic region of *Iss2* (the wild-type strain carrying the chromosome into which the TE had integrated to form TE146) which spans the TE insertion site. Recombinant clones containing the two 12-kb genomic *EcoRI* fragments which comprise this region were isolated from an EMBL-4 library constructed from *Iss2* DNA by virtue of homology with the insert of λ ob 9.04 (see text). The *EcoRI* fragments were subsequently subcloned into pUC9. The extents of λ ob 9.04 and of subclones which were used as probes are shown as solid lines above the restriction map. The TE is represented by the large triangle. Restriction enzyme sites internal to the element were determined from Southern transfer experiments. The site of insertion lies within the 700-bp *HindIII/SalI* fragment and is less than 50 bp from the *SalI* site, the *SstI* site located just within the proximal terminus of the TE being less than 50 bp from this *SalI* site. For this reason, the 700-bp *HindIII/SalI* fragment (probe 2) will only detect fragments spanning the distal junction of the element. R, *EcoRI*; Xh, *XhoI*; Ss, *SstI*; B, *BamHI*; Sa, *SalI*; H, *HindIII*; Xb, *XbaI*. **c**, *Iss2* and TE146 genomic DNA was digested with *HindIII/SalI*, fractionated on a 0.8% agarose gel and transferred to nitrocellulose membrane. The filter was hybridized to probe 4, 32 P-labelled by nick-translation, then washed and autoradiographed. Lane 1, *Iss2*; lane 2, TE146. Label was removed from the filter with two washes in 0.1 M NaOH at room temperature then it was re-hybridized to probe 2, washed and autoradiographed. Lane 3, *Iss2*; lane 4, TE146.

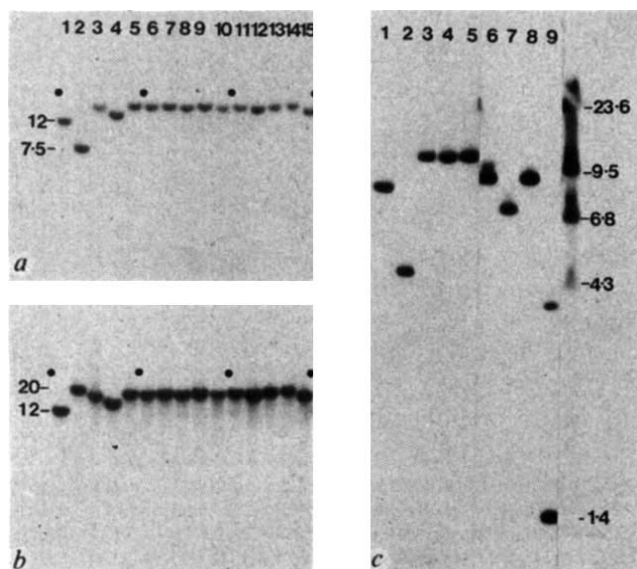
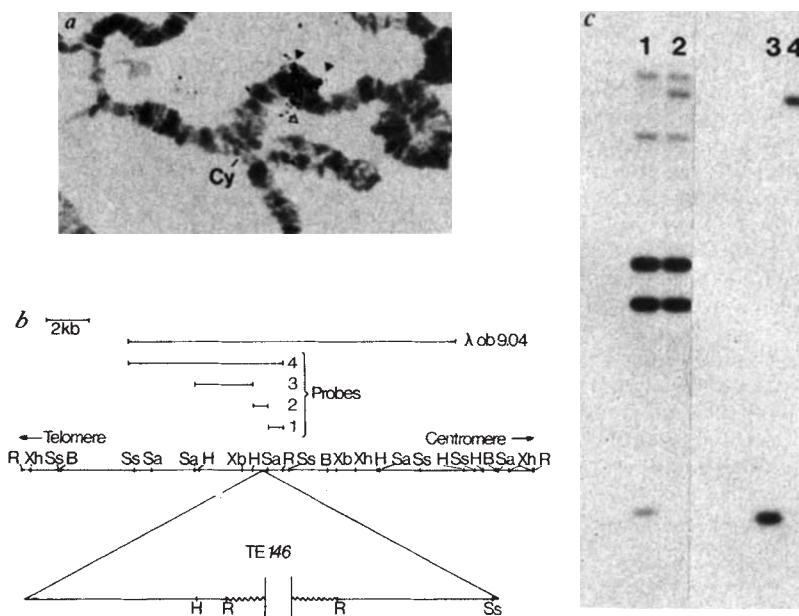


Fig. 3 Restriction enzyme-digested DNA was fractionated on 0.8% (a, b) and 0.6% (c) agarose gels and transferred to nitrocellulose membrane. The filters were hybridized to the probes indicated below, then washed and autoradiographed. Where the same filter was re-used for different probes, label hybridized to the filter was removed by washing with 0.1 M NaOH before hybridization with the subsequent probe. **a**, All samples were digested with *EcoRI*. Lane 1, *Iss2*; lane 2, TE146; lanes 3–15, different strains carrying independent spontaneous excision chromosomes (hybridized with probe 1). **b**, The same filter used for **a** was hybridized to probe 2; an identical result was seen when probe 3 was used. **c**, All samples were digested with *EcoRI/HindIII*. Lanes 1–8, different strains carrying independent spontaneous excision chromosomes; lane 9, *Iss2*. Markers used are *HindIII*-digested λ DNA. Identical results were obtained with either probe 1 or 2. The faint 4-kb band in lane 9 results from partial digestion with *HindIII*. The lanes in **c** and **a** which correspond to the same excision derivatives are as follows: (tracks in **c** are indicated first) 1 = 2, 2 = 4, 3 = 5, 4 = 9, 5 = 8, 6 = 12, 7 = 15; track 8 in **c** is not represented in **a**.

contains two copies of w^+ , was chosen because chromosomes from which this element had been spontaneously lost were available, as was a large region flanking its insertion site. In larval salivary gland chromosomes, this element contains six chromosomal bands, equivalent to several hundreds of kilobases of DNA (see Fig. 1a). Like other members of the TE family, TE146 shares homology with the cloned *w* and FB sequences (Fig. 1b, c). Both copies of the *w* genes carried by TE146 are functional, and spontaneous excision events have previously been isolated by selecting for the absence of w^+ (ref. 7). The site of insertion is associated with *noc*, a gene that is two complementation groups distal to the alcohol dehydrogenase (*Adh*) gene on chromosome arm 2L⁸. Flies carrying homozygous deletions of the *noc* gene are viable, but lack ocelli and associated bristles. TE146 abolishes *noc*⁺ function, so it is associated with a strong recessive *noc* mutation.

The insertion site was sufficiently close to *Adh* to be cloned by chromosomal walking⁹. Using the cloned *Adh* gene as a starting point¹⁰, overlapping DNA fragments were isolated from a wild-type Canton-S library¹¹ until the insertion site was reached. That DNA spanning the insertion site of TE146 had been obtained was shown by *in situ* hybridization to larval salivary gland chromosomes. The phage clone λ ob 9.04, containing DNA located roughly 100 kb distal to *Adh*, hybridizes with both sides of the element (Fig. 2a); it therefore contains DNA sequences spanning the insertion site. Using as a probe the *Drosophila* DNA contained within λ ob 9.04, we isolated a 25-kb segment of DNA from a phage bank constructed from the genomic DNA of the strain of flies carrying the chromosome into which the TE had integrated to form TE146. The restriction map of this DNA, and Southern transfer experiments which localize the insertion site to a 700-base pair (bp) *SalI/HindIII* fragment, are shown in Fig. 2b, c.

We have analysed eight independent spontaneous excision chromosomes derived from TE146 which, in addition to having lost the w^+ markers, are cytologically wild type at the insertion site. Southern filters of gel-fractionated DNA digests from these strains were prepared, and the filters were hybridized to two different probes, one with sequences adjacent and distal to, the other with sequences adjacent and proximal to, the insertion

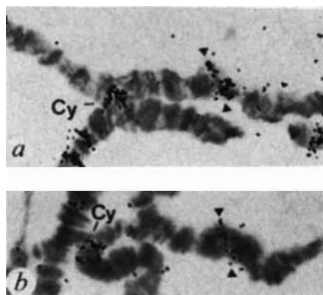


Fig. 4 *In situ* hybridization of the FB8 probe to two different chromosomes that have spontaneously lost TE146 (see Fig. 1c). The arrow indicates FB homology at bands 35B1.2. Cy indicates the left break of *In(2L)Cy*: a, TE146(Z)SW22/*CyO*; b, TE146(Z)SW13/*CyO*.

site (see Fig. 2 legend). Figure 3a, b shows the resulting autoradiograms. In TE146 the distal probe hybridizes to a 20-kb *Eco*RI band which spans the distal junction of the element and adjacent chromosomal DNA, whereas the proximal probe hybridizes to a 7.5-kb band corresponding to the proximal junction. By contrast, each of the lanes containing spontaneous excision chromosome samples contains a single band which appears to be homologous to both probes; this is clearly one band which hybridizes to both probes, and not two different bands of the same size, because consistent results were obtained when this experiment was repeated after the DNA had been cut with different restriction enzymes (data not shown). As homology to both the 700-bp proximal and distal probes is observed, these excision events have not deleted more than 700 bp of chromosomal DNA sequences immediately flanking either side of the insertion site. The simplest interpretation of the data is that these events resulted from the deletion of most, but not all, of the several hundred kilobases of DNA sequence associated with the element. The amount of TE DNA remaining in these chromosomes can be determined from the increased sizes of the wild-type 1.4-kb *Eco*RI/*Hind*III fragment spanning the transposable element insertion site. The fragment varies between 3 and 10 kb in the excision chromosomes examined (Fig. 3c). *In situ* hybridization experiments indicate that the remaining DNA shares homology with cloned FB sequences (see Fig. 4).

FB elements are implicated in mediating these excision events because of their location at the termini of elements belonging to the TE1 family³. The terminal structure of FB elements, combining short tandem repeats within large inverted repeats, could promote homologous pairing, in and out of register, between the inverted repeats⁴. Where two FB elements flank an intermediate sequence, as in members of the TE1 family, pairing and recombination between two repeats (one from each FB element) in the same orientation will result in the loss of the intermediate sequence and the retention of a variable portion of the FB sequences. The loss of DNA flanked by FB elements has been shown previously for two *w* mutations, white-crimson (*w^c*) and Dominant zeste-like (*w^{DZL}*)^{12,13}. There are, however, two differences between the loss of TE146 from *noc* and the losses of the FB-associated insertions from *w*: (1) losses of TE146 are selected on the basis of a function carried by the element itself, rather than a function which is affected by the position of the element; (2) the DNA separating the FB elements is <14 kb for *w^c* and *w^{DZL}*, whereas we estimate a minimum of 200 kb for TE146.

While the events described above, in which essentially all of the TE is lost in a single generation, occur relatively infrequently ($\sim 5 \times 10^{-5}$) (ref. 7), partial deletions of the element, which result in the loss of only one copy of the *w⁺* gene, can occur at much higher frequencies ($\sim 5 \times 10^{-4}$). These partially deleted chromosomes retain several polytene bands associated with the transposable element, and may result from exchange between a

terminal FB element and one internal to the TE, separating its two *w⁺* genes. These partial deletions do not revert *noc*, but subsequent loss of the remaining bands can restore a wild-type *noc⁺* phenotype (D.G. *et al.*, unpublished data). Such losses of TE146 in two stages, via the partial deletion intermediate, result in chromosomes which are indistinguishable from the one-step excision events already described; they retain <10 kb of FB-homologous sequences at the insertion site. The length of the insertion varies among different independently derived losses (data not shown).

Paro *et al.*¹⁴ have suggested that the presence of FB elements at the termini of the TE1 family may be the cause of their genetic instability. The results presented here indicate that the excision of the six intervening chromosomal bands of TE146 from their site of insertion could result from recombination between the terminal FB elements of the transposable element. The imprecise nature of these excisions provides an explanation for the lack of correlation between excision and reversion for most of the mutations induced by this family of elements^{1,2,15}. In the case of TE146, however, all the imprecise excision events, regardless of their mode of origin, restore a wild-type *noc⁺* phenotype. It would be difficult to rationalize this observation if TE146 had disrupted *noc* function by inserting into a coding portion of the gene. A possible explanation of the reversions is that the insertion of several hundred kilobases of DNA in close proximity to *noc* can disrupt its expression by virtue of a position effect. For example, the TE could alter the local chromatin structure and thereby affect *noc* expression. In this case, excision of the element, despite being imprecise, could restore normal expression by removing most of the material responsible for the structural alteration. A similar position effect observed for the *w^{DZL}* mutation is also removed by the deletion of DNA sequences between two FB elements.

This work was supported by MRC grants to M.A. R.K. was an EMBO Fellow and W.C. a MRC Fellow. We thank G. Ising and J. Roote for fly stocks; D. Goldberg, M. Goldberg and S. Potter for recombinant clones; and B. Durrant for drawing the figures.

Received 25 February; accepted 30 April 1985.

1. Ising, G. & Ramel, C. in *The Genetics and Biology of Drosophila* Vol. 1b (eds Ashburner, M. & Novitski, E.) 947-954 (Academic, New York, 1976).
2. Ising, G. & Block, K. *Cold Spring. Harb. Symp. quant. Biol.* **45**, 527-551 (1981).
3. Ising, G. & Block, K. *Molec. gen. Genet.* **196**, 6-16 (1984).
4. Potter, S. S., Truett, M., Phillips, M. & Maher, A. *Cell* **20**, 639-647 (1980).
5. Truett, M. A., Jones, R. S. & Potter, S. S. *Cell* **24**, 753-763 (1981).
6. Goldberg, M. L., Paro, R. & Gehring, W. J. *EMBO J.* **1**, 235-252 (1982).
7. Gubb, D. *et al. Genetics* (submitted).
8. Ashburner, M., Aaron, C. S. & Tsubota, S. *Genetics* **102**, 421-435 (1982).
9. Bender, W., Spierer, P., Spierer, A. & Hogness, D. J. *molec. Biol.* **168**, 17-34 (1983).
10. Goldberg, D. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5794-5798 (1981).
11. Maniatis, T. *et al. Cell* **15**, 687-701 (1978).
12. Collins, M. & Rubin, G. M. *Nature* **308**, 323-327 (1984).
13. Levis, R. & Rubin, G. M. *Cell* **30**, 543-550 (1982).
14. Paro, G., Goldberg, M. L. & Gehring, W. J. *EMBO J.* **2**, 853-860 (1983).
15. Gubb, D., Shelton, M., Roote, J., McGill, S. & Ashburner, M. *Chromosoma* **91**, 54-64 (1984).

Identification of myosin heavy chain in *Saccharomyces cerevisiae*

F. Z. Watts*, D. M. Miller† & E. Orr*

* Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

† MRC Laboratory of Molecular Biology, University Postgraduate Medical School, Hills Road, Cambridge CB2 2QH, UK

Motility in biological systems is expressed in a variety of ways, such as cytoplasmic streaming, cell shaping, nuclear migration and muscle contraction. These functions are thought to be mediated by structural proteins, for example, myosin, actin and tubulin (see ref. 1 for review). The involvement of myosin in muscle contraction is well documented and this protein is implicated in generating the cleavage forces during cytokinesis in some non-muscle cells².

Here, we report the isolation of a protein similar to myosin as judged by its biochemical and immunological properties, from the yeast *Saccharomyces cerevisiae*. Parts of the protein have been conserved through evolution at the protein and DNA sequence levels. The presence of this protein in the region bordering mother cell and bud, as revealed by immunofluorescence, suggests that it is involved in cell division.

In an attempt to isolate novobiocin binding proteins from yeast, we applied total cell extract to a novobiocin-Sepharose column³ and eluted the bound proteins with ATP, urea or increasing salt concentrations. Under these experimental conditions, several proteins were eluted in 1 M KCl or 5 M urea (data not shown). Among these were two proteins of relative molecular mass (M_r) $200,000 \pm 5\%$ present in both fractions. The high M_r of these proteins, combined with the fact that novobiocin inhibits the ATPase activity of bacterial topoisomerase type II through competition with ATP³, led us to examine whether one of these proteins could be a myosin heavy chain, which is known to contain an ATPase activity.

We first used polyclonal⁴ and monoclonal⁵ antibodies raised against nematode muscle myosin, in the Western blot technique⁶. The polyclonal IgG fraction recognizes the smaller of the two high M_r proteins (Fig. 1a, lanes 1, 3). Two monoclonal antibodies (out of a set of 22) raised against nematode myosin interact with the same yeast protein (Fig. 1a, lane 4), supporting this result. One of these antibodies (28.2) interacts with a moderately conserved region of the S2 proteolytic fragment of the nematode myosin heavy-chain isoform encoded by the *unc-54* gene⁵.

Several biochemical criteria have been proposed by Pollard⁷ and others to ensure a positive identification of myosin. The most important of these are its ability to form a complex with actin and its ATPase activity, sometimes stimulated by actin but always dependent on either Mg^{2+} or Ca^{2+} ions. The Ca^{2+} -dependent activity is especially striking, as other known cytoplasmic ATPases are almost exclusively Mg^{2+} -dependent and are usually inhibited by Ca^{2+} . Therefore, this activity is generally used to follow the purification of myosin from crude cell extracts. Table 1 shows that the ATPase activity of the 1 M KCl fraction eluted from the novobiocin-Sepharose column is actin activated (using purified rabbit actin) and is either Mg^{2+} or Ca^{2+} dependent.

To discover whether the putative myosin can form a complex with actin, we isolated actomyosin complexes under conditions similar to those described by Pollard⁷. Yeast spheroplasts were lysed in 10 mM Tris-HCl pH 7.9, 1 mM dithiothreitol; sodium pyrophosphate and sucrose were then added to final concentrations of 50 mM and 0.34 M, respectively, to ensure complete myosin solubilization. After a high-speed clearing spin (30,000 r.p.m.), the supernatant was dialysed against 10 mM sodium citrate, pH 6.5, to precipitate the actomyosin complex. The complex was collected by centrifugation at 30,000 r.p.m.; it is composed primarily of a M_r $200,000 \pm 5\%$ protein (Fig. 1b)

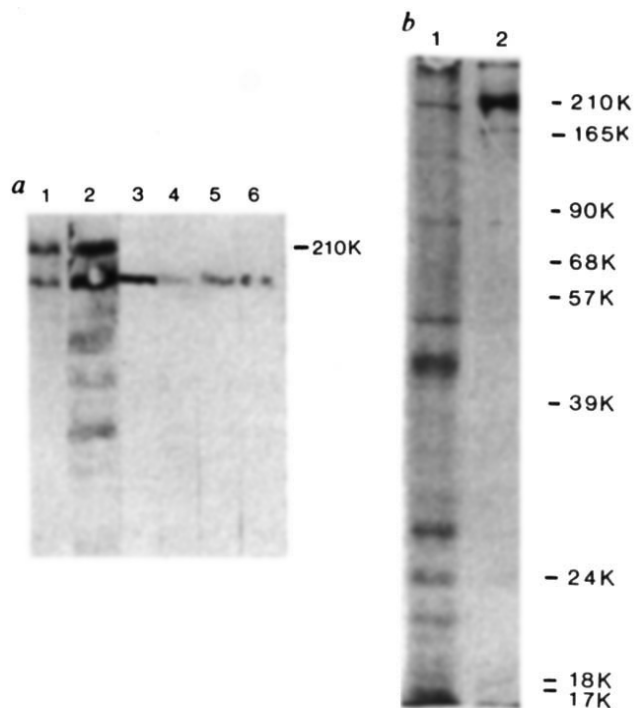


Fig. 1 SDS-polyacrylamide gel electrophoresis and Western blot analysis of crude myosin preparations. **a**, Yeast spheroplasts, prepared using $50 \mu\text{g ml}^{-1}$ zymolyase in 1.2 M sorbitol were lysed in 50 mM Tris-HCl pH 7.9, 1 mM dithiothreitol, 50 mM KCl, 1 mM p-enylmethylsulphonyl fluoride. The resulting solution was subjected to centrifugation at 30,000 r.p.m. for 1 h. The supernatant was loaded onto a novobiocin-Sepharose column, which was then washed with 0.1 M KCl followed by 1 M KCl. Actomyosin complexes were precipitated using a modification of the method described by Pollard⁷. The 1 M KCl fraction (lanes 1–4) and actomyosin complex (lanes 5, 6) were analysed on a 5% SDS-polyacrylamide gel¹⁴ and either (lanes 1 and 5) stained with Coomassie brilliant blue or (lanes 2–4, 6) electroblotted to nitrocellulose⁶. The filter in lane 2 was stained for protein using India ink¹⁵. The filter in lane 3 was incubated with rabbit serum containing polyclonal antibodies against nematode myosin, followed by goat anti-rabbit IgG-peroxidase, while the filters in lanes 4 and 6 were incubated with a monoclonal antibody 28.2 against nematode myosin, followed by goat anti-mouse IgG-peroxidase. M_r marker (210 K) is rabbit myosin heavy chain. **b**, The actomyosin complex was analysed by SDS-polyacrylamide gel electrophoresis (12.5%). Lane 1, yeast actomyosin complex; lane 2, rabbit myosin heavy and light chains. M_r markers (from top to bottom) are rabbit myosin, *Escherichia coli* RNA polymerase, bovine serum albumin and pyruvate kinase.

Table 1 ATPase activity of two crude myosin preparations from *S. cerevisiae*

	Novobiocin-Sepharose 1 M KCl fraction ($\mu\text{mol min}^{-1} \text{mg}^{-1}$)	Actomyosin complex ($\mu\text{mol min}^{-1} \text{mg}^{-1}$)
No additions	≥ 5	≥ 5
Mg^{2+}	150	220
Mg^{2+} /actin	250	220
Ca^{2+}	100	90
K^+ /EDTA	20	10

Assays of ATPase activity were performed in 25- μl aliquots containing 25 mM Tris HCl pH 7.5, 1 mM ATP and one of the following: (1) 5 mM $MgCl_2$; (2) 5 mM $MgCl_2$, 1 mg ml^{-1} rabbit actin; (3) 5 mM $CaCl_2$; (4) 20 mM KCl, 1 mM EDTA. The 1 M KCl fraction had been dialysed to remove the KCl. Inorganic phosphate was measured according to Lanzetta *et al.*¹³. The actin preparation did not contain any detectable ATPase activity.

and actin (as determined with yeast anti-actin IgG in an enzyme-linked immunosorbent assay and on a Western blot, data not shown). Western blot analysis showed that the monoclonal antibodies against nematode myosin reacted with the M_r 200,000 protein (Fig. 1a lanes 5, 6), strongly suggesting the presence of the myosin heavy chain in the complex. The ATPase activity of this complex is again either Mg^{2+} or Ca^{2+} -dependent (Table 1) but is not stimulated by actin, presumably because the latter is already present in sufficient quantities in the complex. In general, myosin is a hexameric molecule composed of two heavy chains and four light chains, each of $M_r \sim 20,000$. Thus far, we have not directly identified myosin light-chain molecules in yeast, although minor proteins of appropriate M_r are co-precipitated in the actomyosin complex (Fig. 1b).

The significance of the immunological data has been further investigated by Southern blot hybridization. A *Bam*HI fragment encoding the ATP-binding region of a nematode myosin heavy chain was hybridized to total genomic yeast DNA. This fragment spans a region that is conserved within the nematode myosin

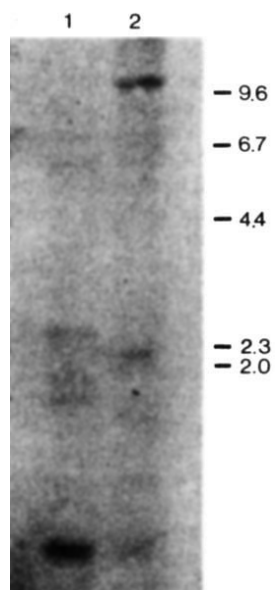


Fig. 2 Southern blot analysis¹⁶ of *S. cerevisiae* genomic DNA with a nematode myosin DNA probe. Genomic DNA from A364A was digested with *Eco*RI (lane 1) or *Bam*HI (lane 2), run on a 0.7% agarose gel and blotted to nitrocellulose. The filter was incubated with a nick-translated *Bam*HI fragment, containing the region of a nematode myosin heavy-chain gene encoding the ATP-binding site⁸, in $2\times$ SSC (0.3 M NaCl, 30 mM sodium citrate) at 60 °C. Size markers are *Hind*III fragments of λ phage DNA.

heavy-chain gene family. Protein sequences in this fragment contain high homology to the ATP-binding region of myosin heavy chains from various organisms⁸. Figure 2 demonstrates that this probe hybridizes to a 1.2-kilobase (kb) *Eco*RI fragment and to a *Bam*HI fragment of ~10 kb. A second DNA probe, carrying a less well-conserved sequence from the S2 region of the nematode *unc-54* myosin heavy-chain gene did not hybridize to yeast DNA (data not shown).

The detection of myosin heavy chain in yeast raises some interesting questions regarding its physiological roles in this and similar organisms. The possible role of myosin in non-muscle cells, as well as that of actin in yeast, have been suggested by their detection using *in situ* immunofluorescent techniques^{2,9,10}. Adopting a similar approach, we used either the nematode anti-myosin IgG or mouse serum containing anti-yeast myosin IgG (our unpublished results). Fixed cells were treated with anti-myosin IgG followed by a secondary IgG coupled to fluorescein isothiocyanate (FITC). Most of the fluorescence was localized at the junction between mother cells and buds (Fig. 3a). This fluorescent band becomes visible as soon as buds can be detected and disappears at the completion of cell division.

These data suggest that myosin is involved in the mechanism of budding, perhaps analogous to its assumed function during cytokinesis in higher eukaryotic cells², which do not have rigid cell walls and in which myosin is associated with the contractile rings of dividing cells. Actin has also been visualized at this neck region in *S. cerevisiae* by the immunofluorescent technique⁹. In this case, however, staining appeared immediately before bud emergence and at the final stage of cytokinesis, but was not detected during bud emergence, presumably because of structural changes in this region during budding. The pattern of staining observed with anti-myosin antibodies coincides with the appearance of 10-nm filaments in this region¹¹. It has been suggested by Kilmartin and Adams⁹ that the actin structures are not associated with the 10-nm filaments in yeast. However, other results (D. Botstein, personal communication) clearly demonstrate that actin is present at the neck region throughout the budding process. We are currently analysing yeast cell-division cycle (*cdc*) mutants that lack the 10-nm filaments at the restrictive temperature¹² to determine whether these filaments contain myosin.

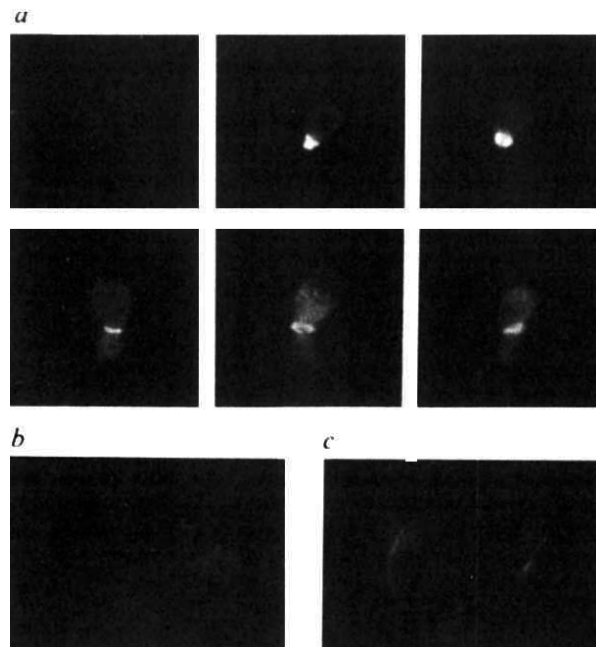


Fig. 3 Fluorescence micrograph of yeast cells treated with anti-nematode myosin antibodies and anti-yeast tubulin antibodies. Immunofluorescence was performed using a modification of the method described by Kilmartin and Adams⁸. Cells were fixed with 3% formaldehyde and spheroplasts were prepared using 100 μ g ml⁻¹ zymolyase. Spheroplasts were incubated with: *a*, serum containing polyclonal anti-nematode myosin antibodies, followed by goat anti-rabbit IgG-FITC; *b*, non-immune serum followed by goat anti-rabbit IgG-FITC; *c*, monoclonal anti-yeast tubulin antibody⁹ followed by anti-rat IgG-FITC.

It has been widely speculated that, as well as its involvement in cytokinesis, myosin is a component of the cellular cytoskeleton. It has been proposed recently that actin participates in the transport and secretion of newly synthesized components for the budding cell wall in yeast^{9,10}. We hope that further analysis of mutants defective in the myosin gene will elucidate its interaction with actin and its function in the cell *in vivo*.

This work was supported by a project grant 8301270 to E.O. from the MRC. We thank M. Pocklington and C. Bagshaw for discussion, T. Harrison and E. Allan for assistance in preparing antibodies J. Karn for the polyclonal nematode antibodies, Ichiro Maruyama for the DNA probes and J. Kilmartin for yeast anti-tubulin and anti-actin antibodies.

Received 5 March; accepted 10 May 1985.

1. Korn, E. D. *Proc. natn. Acad. Sci. U.S.A.* **75**, 588-599 (1978).
2. Fujiwara, K. & Pollard, T. D. *J. Cell Biol.* **71**, 848-875 (1976).
3. Staudenbauer, W. L. & Orr, E. *Nucleic Acids Res.* **9**, 3489-3603 (1981).
4. Macleod, A. R., Karn, J. & Brenner, S. *Nature* **291**, 386-390 (1981).
5. Miller, D. M., Ortiz, I., Berliner, G. C. & Epstein, H. F. *Cell* **34**, 477-490 (1983).
6. Towbin, H., Ramjoue, H.-P., Kuster, H., Liverani, D. & Gordon, J. *J. biol. Chem.* **257**, 12709-12715 (1982).
7. Pollard, T. D. *Meth. Enzym.* **85**, 331-357 (1982).
8. Karn, J., Brenner, S. & Barnett, L. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4253-4257 (1983).
9. Kilmartin, J. V. & Adams, A. E. M. *J. Cell Biol.* **98**, 922-933 (1984).
10. Adams, A. E. M. & Pringle, J. R. *J. Cell Biol.* **98**, 934-945 (1984).
11. Byers, B. & Goetsch, L. *J. Cell Biol.* **69**, 717-721 (1976).
12. Byers, B. & Goetsch, L. *J. Cell Biol.* **70**, 35a (1976).
13. Lanzetta, P. A., Alvarez, L. J., Reinach, P. S. & Candia, O. A. *Analyt. Biochem.* **100**, 95-97 (1979).
14. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
15. Hancock, K. & Tsang, V. C. W. *Analyt. Biochem.* **133**, 157-162 (1983).
16. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).

India-Eurasia collision chronology

Patriat and Achache's analysis¹ of the convergence between India and Eurasia led them to conclude that India collided with Eurasia around anomaly 22 time (~50 Myr), confirming an earlier interpretation based on geological relationships in northwestern India that collision was pre-Middle Eocene², probably around anomaly 23 time³ (~52 Myr). Geological evidence from the Lhasa region^{4,5} also suggests that obduction of the suture-zone rocks onto the leading edge of India occurred there in the pre-Eocene interval, and that any significant shortening in the Lhasa Block was complete before the end of the Palaeocene⁵.

These observations pose a problem for Patriat and Achache's model¹ because in their Figs 1b and 4, 50 Myr ago southern Tibet was located some 7° north of the presently exposed northern boundary of India. Patriat and Achache¹ suggest that the 700 km of continental crust necessary to fill this gap was consumed first by subduction of ~400 km of Indian continent beneath southern Tibet from 50 to 44 Myr, and then by continued continental underthrusting and internal deformation to account for another 300 km of north-south shortening before 37 Myr. There are at least three problems with this model.

(1) Deformation in the supposed continental subduction and underthrust zone (Indus-Tsangpo suture) should be of Eocene age, rather than late Cretaceous or early Palaeocene as shown by field relationships^{4,5}.

(2) The now-vanished continental crust would have lain between the Tethyan Himalaya and southern Tibet, in which case the Tethyan Himalaya would have lain in the interior of the Indian continental block before collision, and not at its leading edge as is widely accepted^{2,6,7}.

(3) Southern Tibet would have had a much thicker than normal continental crust from mid-Eocene to present, in which case isostatic rebound would have led to a Palaeogene rise of the Tibetan plateau, for which there is no geological evidence.

These problems disappear if the pre-collisional shape of India is extended northwards to take account of its original size in Gondwanaland^{8,9}. A pre-collisional Greater India with its northern edge extending as far north of the exposed Indian Shield as the Kun Lun is today would not only have collided with southern Tibet by the postulated 50-Myr collision time, but could later have underthrust its broken-off leading edge, the Tethyan Himalaya, to produce the 70-km continental crust beneath Tibet. The underthrusting could have been as a series

of intracontinental slices¹⁰ rather than a single continental underthrust^{2,3}.

Recognition of the Tethyan Himalaya as the leading edge of Greater India obviates the necessity to postulate 700 km of Eocene continental subduction and underthrusting beneath southern Tibet, and raises the possibility that the threefold sequence of events suggested by Patriat and Achache occurred in almost exactly reverse order. Initial shortening within Eurasia is taken up by block rotation and lateral extrusion along sinistral strike-slip faults¹¹, then followed by crustal thickening through internal deformation and the beginning of continental underthrusting, and finally, when the continental underthrust zone is sufficiently well established, large-scale continental subduction occurs^{2,9}.

C. MCA. POWELL

P. J. CONAGHAN

*School of Earth Sciences,
Macquarie University,
North Ryde, New South Wales 2113,
Australia*

C. T. KLOOTWIJK

*Bureau of Mineral Resources,
Geology and Geophysics,
PO Box 378,
Canberra City, ACT 2601,
Australia*

1. Patriat, P. & Achache, J. *Nature* **311**, 615-621 (1984).
2. Powell, C. McA. & Conaghan, P. J. *Earth planet. Sci. Lett.* **20**, 1-12 (1973).
3. Powell, C. McA. & Conaghan, P. J. *Geology* **4**, 727-731 (1975).
4. Burg, J. P., Proust, F., Tapponnier, P. & Chen Guo Ming. *Ecol. geol. Helv.* **76**, 643-665 (1983).
5. Burg, J. P. & Chen, G. M. *Nature* **311**, 219-223 (1984).
6. Allegre, C. J. *et al. Nature* **307**, 17-22 (1984).
7. Andrews-Speed, C. P. & Brookfield, M. E. *Tectonophysics* **82**, 253-275 (1982).
8. Veevers, J. J., Powell, C. McA. & Johnson, B. D. *Earth planet. Sci. Lett.* **25**, 383-387 (1975).
9. Powell, C. McA. in *Pakistan Geodynamics Project* (eds Farah, A. & de Jong, K. A.) 5-24 (Geological Survey of Pakistan, Quetta, 1979).
10. Hirn, A. *et al. Nature* **307**, 25-27 (1984).
11. Tapponnier, P., Peltzer, G., Le Dain, A. Y., Armijo, R. & Cobbold, P. *Geology* **10**, 611-616 (1982).

ACHACHE AND PATRIAT REPLY—Powell *et al.* estimate the pre-collisional shape of India after 'its original size in Gondwanaland'. This coincides with the present position of the Kun Lun belt. We think this method is highly questionable and thus, such an estimate must be considered as speculative.

The problems allegedly raised by our model call for the following comments: (1) The Cretaceous to Palaeocene F₁ deformation observed in the Yarlung Zangbo suture zone is related to northward subduction of oceanic crust^{1,2}. It is followed by a larger-scale although less intense younger F₂ deformation which can be associated with crustal subduction and shortening^{1,2}.

(2) The observation of Tethyan sediments in no way contradicts the model of subduction of 400 km of continental crust and only requires a 500-km-wide continental shelf north of India, before the collision. Whether this can still be considered as the so-called leading edge of India is merely a matter of vocabulary.

(3) The age of rise of the Tibetan plateau is still an open question. In particular, there is no valid evidence against a Palaeogene rebound.

We therefore do not see any real objection to our model in the comment by Powell *et al.* Furthermore, several field observations seem to contradict the reversed sequence of events suggested by these authors. First, a recent palaeomagnetic study³ of Tethyan sediments in Tibet has shown the amount of shortening in the MCT to be 400 ± 400 km, thus much less than required by the proposed model. Second, structural and geochronological data indicate that significant underthrusting occurred in the Indian continent (Kangmar Thrust, Main Central Thrust, Main Boundary Thrust) between 45 and 35 Myr ago (see ref. 3). And third, the magnetic anomaly pattern in the South China Sea indicates that the motion along the Red River fault (and thus the extrusion of Indochina⁴) did not start until anomaly 13 time (36 Myr), that is, 15 Myr after the beginning of the collision (see ref. 5).

J. ACHACHE

P. PATRIAT

*Institut de Physique du Globe de Paris,
Laboratoire de Géomagnétisme et
Paléomagnétisme,
4 Place Jussieu, 75005 Paris, France*

1. Burg, J. P., Proust, F., Tapponnier, P. & Chen, G. M. *Ecol. geol. Helv.* **76**, 643-665 (1983).
2. Burg, J. P. & Chen, G. M. *Nature* **311**, 219-223 (1984).
3. Besse, J., Courtillot, V., Pozzi, J. P., Westphal, M. & Zhou, Y. X. *Nature* **311**, 621-626 (1984).
4. Tapponnier, P., Peltzer, G., Le Dain, A. Y., Armijo, R. & Cobbold, P. *Geology* **10**, 611-616 (1982).
5. Patriat, P. & Achache, J. *Nature* **311**, 615-621 (1984).

Matters Arising

Matters Arising is meant as a vehicle for comment and discussion about papers that appear in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted. Neither contribution nor reply (if one is necessary) should be longer than 500 words and the briefest of replies, to the effect that a point is taken, should be considered.

Image analysis takes off

from Geoff Jenkinson

Recent software developments are taking image analysis into new and uncharted areas of science.

THE basis of the image analysis method is well-known and it is not intended to describe this here. However, the architecture of the system is relevant in defining the capabilities of an image analysis system. A modern instrument, for example the Quantimet 920, will consist of a number of integrated components as shown in Fig. 1.

The video processor takes in the analogue video signal from the scanning source, either a TV-type scanner or a scanning electron microscope. The scan standards are chosen to optimize the resolution and signal potential of the source and give over 630,000 image points with up to 256 grey levels. A built-in shading corrector compensates for scanner and optical grey level variation. The resulting signal is converted to digital form and can be stored as a full grey-level image.

The 2-D array processor performs real-time image transformations to remove errors caused by optical and electronic effects which limit the rise-time of the video signal, to enhance fine detail, for example in grain boundaries, and to remove specimen-dependent shading. The image processor performs other digital signal processing tasks and also extracts a wide range of field and feature data parameters from the image. The image store can store multiple grey images at a variety of resolutions and binary images. These images may be roamed and zoomed and transferred to and from disk.

Software

Users of image analysis systems have an enormous variety of requirements. The software has to be structured to meet all these needs and can be broadly considered

Fig. 2 Image analysers can now be packaged together with a scanning electron microscope, as on this Quantiscan 922 from Cambridge Instruments.



as falling into three main types:

Turnkey application program. A program or set of programs is provided which requires only that the operator follows a sequence of prompts which appear on the display. This will be normally used for routine applications which only require the operator to make minor adjustments to settings of detection or for interactive image editing for example.

Interactive routine building. To create a new application solution the user is provided with a special ready-to-use interactive software package known as QUIPS. This allows the various image analysis facilities to be called up and the effect of selecting various combinations made immediately visible. Test results may be examined to decide if it is worthwhile proceeding to a more detailed analysis. In operation the primary image analysis functions are displayed on the control monitor and these are selected by means of a pointing finger which is manipulated via a digitizing tablet.

Software extensions. The potential of the system is not limited to the standard functions as described above. A PASCAL software package allows complete freedom of access to all data and images in the system, so that new functions and instructions can be defined and incorporated into the standard routines.

Application examples

Two applications have been chosen which illustrate the range of techniques required of image analysis systems. These are all in routine use on standard production instruments.

Receptor binding autoradiographs. In quantitative autoradiography in neurobiology and pharmacological research, image analysis provides facilities for region selection of the autoradiography, measuring radiation levels and colour-coding images. Image analysis can significantly

cantly ease the interpretation of results and by means of its speed, make possible experiments which could not previously be contemplated.

In the study of receptor binding the general approach can be illustrated by quantifying the binding of a drug to rat forebrain sections. Adjacent sections are incubated both with and without a drug which displaces a radioligand from the binding site or sites which are being investigated. Thin sections of brain tissue are used to expose a tritium-sensitive film which produces an autoradiograph in which the radiation levels give a corresponding density on the film. Standards of radio-labelled tissue are also exposed on the film. A calibration curve is produced which converts the grey-level on the film to the quantity of the bound drug.

The adjacent sections produce images which represent total binding and non-specific binding of the drug. A single section of total binding is imaged at low magnification onto the TV camera of the image analyser and stored in its image memory. Corrections of non-uniform illumination and conversion to radiation via the calibration curve are made before storage. The image of the nonspecifically bound section is superimposed with the outline of the previous section and similarly stored. Subtraction of the nonspecific bound image from the total bound image now gives an image which represents the specifically bound component. The percent specific binding can now be determined by dividing the specific bound image by the totally bound image. Regions of interest may be outlined by means of the digitizing tablet and the mol per mg protein within the region determined.

An alternative method of measuring the radiation values is to use a higher magnification and dark-field illumination on a light microscope which allows the individual grains in the photographic emul-

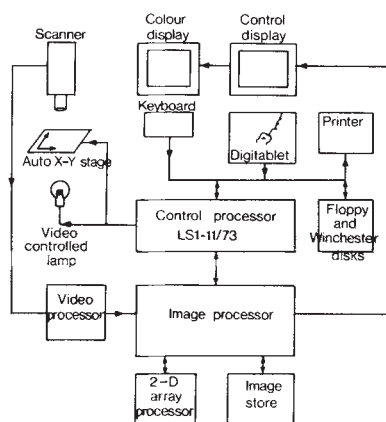


Fig. 1 Block diagram of a modern image analysis system.

sion to be seen and hence counted. This method is used when smaller regions need to be identified or when the density of the autoradiograph is low.

Combined scanning electron microscopy and image analysis. The potential of combined scanning electron microscopes and image analysis systems has been acknowledged for some years. The SEM gives superior resolution and excellent contrast on many specimens which cannot be analysed by means of a light microscope. A variety of detectors is available for the SEM which form images corresponding to different properties of the specimen and which are suitable for image analysis. Up to now however it has been difficult to capitalize on this potential due to the relative complexities involved in connecting the two instruments together. It is now possible to fully integrate the capabilities of the electron optics and image analysers, as is done in the Quantiscan 922 system shown in Fig. 2. This analyser has a resolution better than 4nm in secondary electron imaging mode when fitted with a directly heated LaB₆ emitter producing an overall magnification range of 4× to 300,000×.

The SEM image is transferred directly to the image store in the image analysis system from where it can be continuously displayed on the screen of the analyser. Transfer rate is selectable from 0.3s up to 200s, and calibration of the measurements is made automatic by monitoring the SEM operating conditions.

When used in conjunction with an energy dispersive X-ray system the analytical potential is further extended. The X-ray information indicates the presence of particular elements in a specimen, and the importance of this is that it may be combined with the topographic and structural information from the image analysis to give a complete interpretation of the specimen.

Other applications

Image analysers can be used for many other applications and have found ready acceptance in quality control in fibre industries where they are used to measure the characteristics of synthetic fibres, in electronics to map the dislocation density in semiconductor wafers and to monitor oil contamination in hydraulic systems. In the field of metallurgy, where image analysers have been in use for more than two decades, their utility is being extended by the introduction of automatic chart-rating.

There is also increasing demand for fast simple image analysers capable of routine particle sizing and counting and these are incorporating many of the features of their larger relatives. This will continue the downwards cost trend which will make image analysis available to an even greater range of users in the future. □

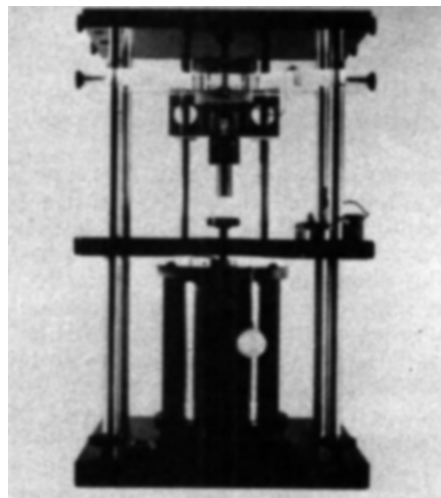
Geoff Jenkinson is at Cambridge Instruments Ltd, Rustat Road, Cambridge CB1 3QH, UK.

Image analysis and microscopy

A sampling of products from a simple magnifier to the latest in image analysis.

- Capable of analysing images from microscopes, video-tape recorders, drum scanners, photometers and photographs, Reichert-Jung's Polyscan offers maximum flexibility of image format without compromising access to the powerful software library. The system is programmable using either direct routines from an extensive menu or custom modifications entered via the image processing library. Polyscan will run automatically and uses a high resolution, low geometric distortion camera, capable of better than 1,000 lines, and computer-controlled beam blanking for image integration. Pseudocolour processing is supplied as standard and real colour is available as an option. Two CRTs are included in the system: a 12-inch black and white VDU for displaying user dialogue and programming; a 13-inch high resolution colour monitor for the image being analysed.

Reader Service No. 100.



Gentleman Jim for rapid freezing.

- A novel method for quick freezing tissue samples for electron microscopy is adopted in a new system from Pelco. 'Gentleman Jim' is an instrument which permits instant freezing of biological samples by dropping the selected tissue onto a polished surface of pure copper at liquid nitrogen temperature. Gentleman Jim does not permit bounce during contact as a special hydraulic system forces the sample tissue to maintain contact during the crucial initial moments of quick freezing. Excellent results have been obtained for the first 12 to 15 micrometres of tissue depth. Helium is unnecessary since liquid nitrogen produces satisfactory results. After freezing, dehydration by freeze substitution may be carried out.

Reader Service No. 101.

These notes are based on information provided by the manufacturers. To obtain further details about these products use the reader service card bound inside the journal.

- The DM54K image analyser from Digital Measurement Associates Ltd. provides a high speed low cost approach to obtaining quantitative measurements from subjective images. The system comprises a television camera interfaced to an Apple IIe computer via a custom interface enabling video images to be digitized in real time, approximately one fiftieth of a second. The interface expands the Apple working memory to 128K and also provides eighty-column text facility. It is supported by a suite of dedicated software providing a full range of image analysis features. The TV camera may be attached to a standard viewing system to image macro objects, photographic prints and so on. Alternatively a standard range of adaptors is available for light microscopes. TV images are digitized in real time to provide a detected matrix of 280 × 192 pixels. A possible 256 grey levels are available in either 'blackier than', 'whiter than' or 'grey slice' modes together with a programmable shading correction system to adjust for uneven background illumination. Area, end count, full feature count and intercept data are generated as a basis for a full stereological analysis.

Reader Service No. 102.

- The new 40-10 model image analyser from AMS (Analytical Measuring Systems) is available with a dedicated biological software package for applications such as colony counting, bacteria counting, breakpoint MIC determination, inhibition zone measurement, viral plaque counting and spiral plate counting. The VIDS II semi-automatic image analyser is a graphics tablet based system and is used in conjunction with an Apple computer and Epson printer. The scanner provided with the system can be mounted on any type of microscope or used with the AMS macro stand for macro-samples. With this system any sample that can be imaged with the camera can be "drawn round" and measured. Typical applications include analysis of histological sections, SEM and TEM micrograph measurements and particle size analysis.

Reader Service No. 103.

The AMS 40-10 image analyser.





A Nimbus microcomputer forms the basis for the latest image analysis package from Digithurst.

• As an alternative to supplying image analysis equipment for connection to existing microcomputers, Digithurst has introduced an image analysis system based around the Research Machine's Nimbus. The latest version of the Digithurst Micro-Sight is incorporated in the system which enables images to be captured at up to 512×512 resolution with 265 grey levels. Using standard graphics the image can be displayed at 256×256 with 16 grey levels on the Nimbus screen. The system uses an 80186 CPU running at 8MHz 320K of memory and twin floppy disk drives; hard disk drive versions are available together with networking options. The memory can be expanded up to 1 megabyte. Images captured by the system can be dumped to a printer or stored onto disk, they are also available to graphics packages which can manipulate and edit the displayed pictures in applications such as art, design, advertising, display and computer aided learning.

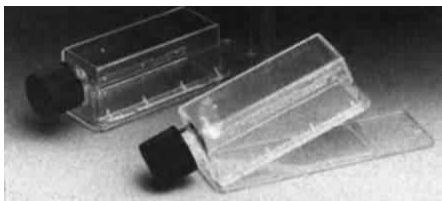
Reader Service No. 104.

• Raymond A. Lamb's Labstack filing system offers a wide range of stackable steel cabinets for storage of microscope slides, either vertically or horizontally in slotted racks. Transparency cabinets are also available as well as paraffin block filing cabinets with divided drawers to hold either System I embedding rings or System II embedding cassettes. These cabinets are in two sizes, either standard for floor stacking or half-depth for bench-top use. The drawers are provided with stops to prevent accidental removal although they can be removed and relocated. Accessories include spring spacers for initial filing before the mountant has hardened for contact storage, adaptors for holding larger slides in 26 mm wide drawers, different coloured location slips, transparency filing cards, slide wallets and dividers to file 76×26 mm slides in 52 mm wide drawers. System I embedding rings and System II embedding cassettes are included in the range together with cardboard boxes in varying sizes for unmounted blocks.

Reader Service No. 105.

• Philips Analytical has introduced a new Edax EDSCAN software package which provides a means of acquiring and presenting spatially related chemical information from a sample using a scanning electron microscope with attached Edax 9100 energy dispersive analyser. The package uses computer controlled microscope beam positioning to probe the specimen and collect from it X-ray and electron back-scattered signals quickly and efficiently. Collection beam geometries range from single spots to digitized line scans and area scans. Data is stored on disk and can, if required, be easily manipulated using simple arithmetic processing before presentation on one of a variety of available output devices. The Fortran-based EDSCAN software package is divided into two distinct functions: XSCN, devoted to X-ray signal collection, and ESCN for microscope signal acquisition.

Reader Service No. 106.



Nunc's flask-come-slide.

• The Nunc SlideFlask is a combination of a cell culture flask and a microscopy slide. The bottom of the flask has exactly the same dimensions as a standard microscopy slide, giving a culture area of 9 cm^2 . It is manufactured from Nuncolon Delta material for good cell adherence and optical clarity. The two flask parts are welded together, but slide and flask can be separated at any step in the preparation procedure after the cells have been cultured. The Nunc SlideFlask is specially designed for culture and preparation of cells and subsequent microscopic examination of the specimen under high power magnification (for applications such as chromosome examination, single cell autoradiography and single cell immunofluorescence).

Reader Service No. 107.

ADVERTISEMENTS

Automatic Tracking

Track that rat, insect, fish, limb, vehicle, crack, etc.

Tracking unit accepts input from video camera or recorder, and outputs the location of a high-contrast object in the TV image.

Can be interfaced to Apple II, BBC, etc.

HVS IMAGE ANALYSING LTD,
22 Cromwell Road, Kingston,
Surrey, England KT2 6RE
Tel: 01-546 9090



Reader Service
No.43

PHORBOL ESTERS

CALL OR WRITE
FOR
FREE CATALOG



LC SERVICES CORPORATION

165 New Boston Street
Woburn, MA 01801
(617) 938-1700

Reader Service No.44

SINGER MICROMANIPULATOR MK III Patented



Perfect control in 3 Dimensions from a single joystick. Continuous variable reductions to suit all magnifications. No after sales maintenance needed.

Details from:
Singer Instrument Co. Ltd.,
Treborough Lodge Roadwater,
Watchet Somerset TA23 0QL
England. Tel. (0984) 40226 Telex:
337492 Comcab G.

Reader Service No.45

nature is available in microform from University Microfilms International.

☐ Please send information about these titles:

Name _____

Company/Institution _____

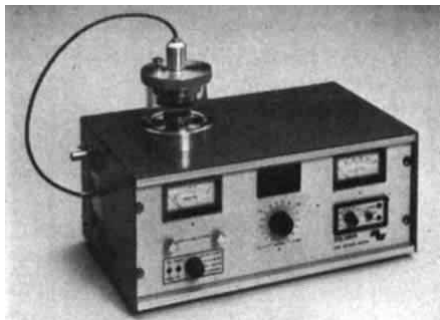
Address _____

City _____

State _____ Zip _____

Phone () _____

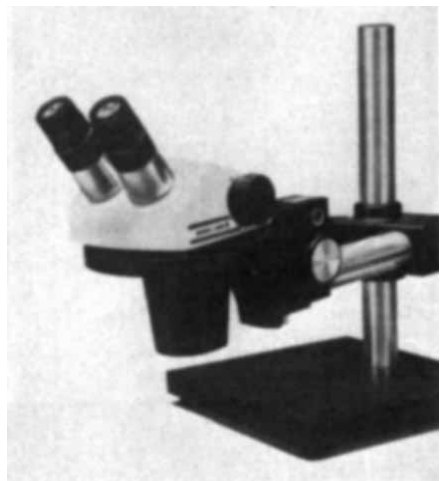
Call toll-free 800-521-3044. Or mail inquiry to:
University Microfilms International, 300 North
Zeeb Road, Ann Arbor, MI 48106.



Polaron's new splutterer.

● Although sputter coaters are now in routine use in the scanning electron microscopy laboratory for the coating of biological and other types of insulating materials with gold or some other noble metal, it has hitherto proved impossible to sputter materials such as aluminium or carbon which are necessary as a conducting coating for microprobe techniques. But the Polaron E5800 plasma enhanced a.c. sputtering system successfully deposits aluminium and other refractory materials. The plasma in the vicinity of the sputtering target is enhanced by an intense magnetic field; sputtering voltage is typically 300V, 150mA. Sputtering rates for aluminium are typically 60nm per second. If a gold target is used for routine SEM coating, then coating times can be as short as 15–20 seconds, avoiding heating problems on the sample surface.

Reader Service No. 108.



StereoZoom 6 from Bausch & Lomb.

● StereoZoom 6, the latest microscope from Bausch & Lomb, has improved optics designed to match the colour correction response of the human eye. It provides clear, crisp images, true colour rendition and sharp contrast. Standard features include a one-to-six zoom system with an 0.67 to four basic magnification range, sharp resolving power at higher magnification, a wide field of view and exceptional depth of field at lower magnification (0.55 mm at 1.0×). The instrument's InstaZoom makes it possible to run through the entire zoom range with less than one-third of a revolution of the control, and its narrow, 56mm objective end allows it to be used in tight spots.

Reader Service No. 109.

● LKB-Produkter AB of Sweden has introduced a new high-precision glass-knifemaker (LKB 2178 KnifeMaker II) which allows the user to prepare 45° knives in either of two ways: in the traditional way or by applying the new 'balanced break' concept. No special skills are required to produce knives for semi-thin light microscopy sectioning or for cutting ultrathin sections for ultramicrotomy and cryo-ultramicrotomy. In the balanced break technique a standard 400-mm long strip of glass in any of three thicknesses (6.4, 8 or 10mm) is continuously broken up into equal halves to produce a full set of flat-sided 25×25 mm squares. The KnifeMaker II introduces two new scoring positions, to give four scoring choices. By enabling optimal corner-to-corner breaks, it simplifies the production of 45° angled cryo-knives. The holder also has adjustable click-steps, useful when preparing 40, 50, 55 and 60° knives.

Reader Service No. 110.



Bench-top freeze drying by FTS.

● The Tis-U-Dry is a bench-top freeze dryer which allows the histologist to freeze dry and fix heat-sensitive tissues in one chamber. The Tis-U-Dry series has condenser configurations to -130°C and a temperature controllable tissue platten to -100°C . A formaldehyde heater is connected directly to the freeze dry chamber and an optional resin embedding accessory that is mounted in a vacuum top plate allows all steps to be performed directly in the Tis-U-Dry, which is produced by FTS Systems Inc.

Reader Service No. 111.

● New from Paar Scientific is an economical microscope warm-stage which can be fitted easily to virtually any model of microscope. The device consists of the stage itself and a control box on which any temperature from ambient up to 60°C can be set quickly and easily to appear on a digital display. The stage measures 75 mm by 50 mm with a 6 mm aperture and the sample or specimen is protected from induced currents by a bifilar heating element specially constructed of a foil material which covers the entire stage area to prevent cold spots.

Reader Service No. 112.



The Olympus IMT-2.

● New from Olympus is the IMT-2 inverted research microscope specially designed for the examination of cell cultures. Attachments can be fitted to the IMT-2 for brightfield, phase contrast, differential interference contrast and fluorescent observations. Switch-over between these different observation methods is fast and simple, and a fixed stage system assures reliability during micro-manipulation. The microscope features an automatic exposure 35 mm camera with the facility to accept further camera attachments.

Reader Service No. 113.

● Two recent publications by Olympus Optical are available free of charge from Gallenkamp, the company's UK agents. The first, *Photomicrography*, is intended for scientists who have never taken a photograph down a microscope but also for experienced photomicroscopists who would like to improve their techniques. The second booklet, *Fluorescence Techniques in Microscopy* covers selection of fluorescence equipment, filters and methodology. Write to Olympus Department, Gallenkamp, Belton Road West, Loughborough LE11 0TR, stating which booklet is required.

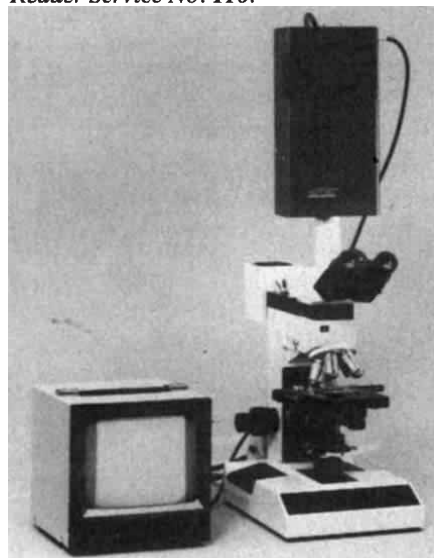
Reader Service No. 114.



The Nikon Microphot-PX is equipped with a sophisticated built-in automatic exposure photomicrographic system, and a wide range of functions for quantitative microscopy. Use the Reader Service Card (no.115) for full details.

• The Carl Zeiss Jena range of stereomicroscopes now includes the GSM, a light but robust stereomicroscope of high optical performance constructed following the Greenough principle. A twin lens slide allows rapid change of magnification and the extensive range of accessories allows various modes of illumination and measurement. The Zeiss Technival 2 is a high performance stereomicroscope with a five step magnification changer giving a 10:1 overall factor. It is a modular system giving a high degree of flexibility in magnification, lighting and photographic attachments. Most of the accessories for this instrument also suit the Citoval 2, a zoom stereomicroscope with a range of 10:1.

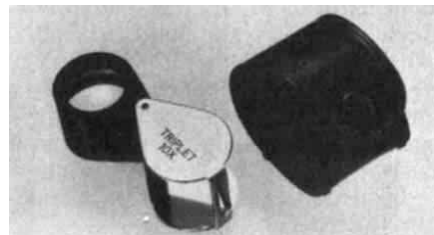
Reader Service No. 116.



Artek's new camera and monitor.

• Artek Systems Corporation has developed a 700-line camera and 700-line monitor for image analysis, dimensional measurement, particle analysis and visual inspection. A wide selection of image tubes is available for a variety of applications including easy viewing of microscopic samples. The camera and monitor system are compatible with Artek's image analysers and new Video Micrometer and Reticle Generator.

Reader Service No. 117.



Pocket magnification from Microtec.

• An extremely highly corrected 10× pocket magnifier has been added to the Microtec range of magnifiers. The lens construction is a new design which enables the high correction to be achieved with lenses over 20mm in diameter. The magnifier has a chromed brass frame and is supplied complete with a hide case.

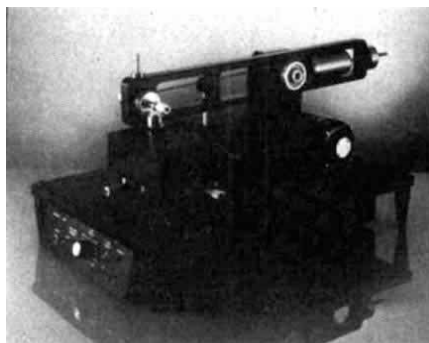
Reader Service No. 118.



The Vias workstation complete with digitizing tablet.

• Vias is a complete video image analysis workstation which includes both a flexible statistical package and a digitizing tablet. Featuring an easily accessible, completely menu-driven computer program, this system quickly and easily summarizes and performs complex morphometric calculations in any scale or coordinate system. Vias is available from Ted Pella Inc. of Tustin, California.

Reader Service No. 119.



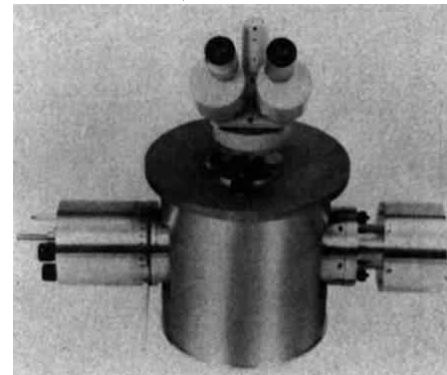
TEM preparation by 'dimpling'.

• The 'Dimpler' is a new prethinning instrument for TEM specimen preparation. Taking a 2.3mm or 3mm diameter specimen with a thickness of 100 micrometres or more, the VCR Dimpler D500 will delicately grind and polish a 'dimple' into the centre of the specimen, leaving a large area which is close to electron transparency. The perimeter of the disk is polished flat but remains in the form of a thicker rim so the specimen can be handled with more stability and greatly reduced risk of damage. It is common to achieve a 95% successful specimen yield. A variety of materials can be 'dimpled' — from the hard ceramics, glasses, quartz, sapphire and silicon, to the soft in v semiconductor materials such as gallium arsenide or indium phosphide. The Dimpler Model D500 is available in Europe through Testbourne Ltd.

Reader Service No. 120.

• The Ion Tech FAB 306A fast atom bombardment specimen milling module extends the scope for specimen preparation techniques for both transmission and scanning electron microscope work. As well as milling conventional materials such as metals, ceramics and glasses, this new FAB equipped atom milling module enables more exotic materials to be thinned safely, thus extending the process into the latest semiconductor, polymer and biological technologies. Materials such as gallium arsenide, silicon oxides and many other very sensitive specimens — even delicate biological structures such as brain cells — can be processed in this new instrument. The FAB 306A combines improved wide dynamic range fast atom guns which produce beams from very low to high current density equivalents of charge free atoms with an efficient liquid nitrogen cooled specimen stage. Both charge and radiation damage are thus greatly reduced and a sensitive automatic terminator allows the milling process to be stopped instantly as desired. A wide beam atom gun facility also enables large area SEM specimens to be surface etched and removes smears and debris caused by mechanical polishing or other physical or chemical preparation techniques.

Reader Service No. 121.



TEM and SEM preparation using fast atoms.

Education's response to adversity

from Richard Pearson

The small-print in the UK government's recent review of the future of higher education is worthy of close examination.

THE UK government's long awaited 'Green Paper', *The Development of Higher Education into the 1990s*, has been widely criticized for its lack of originality and new ideas. But students, educationalists and employers alike will ignore it at their peril. If the final policy statement appearing in 1986 after the period of consultation follows the expected pattern, then many students may have their higher education aspirations frustrated, teachers will have increased workloads and decreased job opportunities, and while many employers will gain from the likely increase in the output of industrially relevant graduates, they may face a growing shortage of good generalists. Driven by the twin pressures of financial restraint and the demographic change which will significantly reduce the 18-year-old cohort over the next ten years, the paper sets out the government's expected priorities for higher education to the end of the century, a key priority being to make higher education more responsive to the needs of the economy.

It is important to get the demographic influence in perspective. There are two features that everyone agrees on. First, the major source of entrants to higher education will continue to be the 18–21-year-old population, and second, the size of this cohort will fall by one-third over the next decade, with most of this fall coming in the early 1990s.

In 1983 the Department of Education and Science (DES) published details of its then current thinking on the future demand for higher education. The report led to widespread comment and criticism from, amongst others, the Association of University Teachers, the Royal Society, the Royal Statistical Society and the Committee of Vice-Chancellors and Principals, who all agreed that DES had taken insufficient account of the rising proportion of women and mature entrants wishing to enter higher education, and the social class differential in the changing birth rate. In particular, it was noted that the birth rate for social classes I and II, which has provided nearly two-thirds of university entrants in the past, rose during much of the relevant period and as such the overall fall in student demand would not be as great as simple demographic trends suggested. There was also criticism that that no allowance had been made for policies that might increase the participation from working class families, or improve accessibility for mature students.

In the light of these criticisms, the DES forecasts were revised in 1984. The key figures are that the number of 18-year-olds in the population will fall from 914,000 in 1984 to a trough of 613,000 in 1995 before rising to 681,000 in the year 2000. Two projections of qualified demand are given by DES (Fig. 1). The projection, Variant X, assumes that student demand will continue at the levels prevailing in 1981–82, and that women will continue to catch up with men in their degree of participation in higher education. DES assumes, however, that the qualified demand index for women will

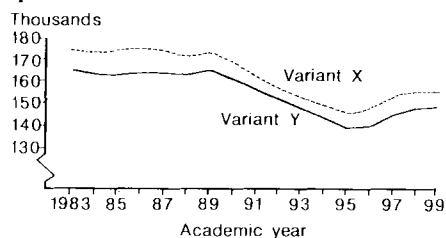


Fig. 1 Projections, from the DES report, of the total number of home initial entrants for the academic years 1983/4 to 1999/2000. Variant Y is the revised version.

not exceed 85 per cent of the observed level for men.

Variant Y, revised again in January 1985, assumes that student demand will continue at the levels observed in 1983, after the public expenditure cuts had caused some redistribution of higher education places. Both variants are based on assumptions about the prevailing real value of student grants and the influence of various social and economic circumstances, including labour market opportunities for graduates. DES concedes that, by 1999, actual student demand could be outside the limits set by Variants X and Y. The figures suggest a fall in student demand over the period to 1995 of about one-fifth.

For young home initial entrants to higher education (aged under 21 years), Variant X shows demand rising slightly to a peak of 134,500 in 1985 before falling back to 107,500 in 1994 and rising again to 117,700 in 1999. It is assumed here that the age participation index (API) will rise from 14.4 per cent of the age group to 16.9 per cent by 1999. Variant Y anticipated that young entrant demand had peaked in 1983–8, and that the peak age participation index assumed (for 1999) is 13.9 per cent of the age group.

Mature entrants have to be added to these totals, and the DES projections pro-

vide for a small growth from 37,500 in 1983–84 to about 40,000–43,000 mature entrants for the academic year 1989–90 and about the same number in 1994.

The government's current expenditure plans assume that expenditure will be available to meet a level of demand equal to variant Y. The University Grants Committee, interpreting these figures, has said that universities should plan on the assumption of a 2 per cent per annum cut in resources until 1990. This could be equivalent to the closure of a major institution each year.

A key theme of the green paper is meeting the needs of the economy which is seen to require a switch in places towards science and technology subjects. There is, however, some doubt whether suitably qualified and interested students will exist to fill the additional places. The paper finds it remarkable that the public sector institutions cannot fill their current places in science and technology subjects. Student interest in science and technology will therefore have to increase significantly over the next few years if it is to counteract the downward demographic trends and allow for the increased number of places to be filled.

Perhaps the most profound and unexpected part of the green paper is the appendix in which performance measures are proposed to help evaluate the effectiveness of individual institutions. Costs per student, years per graduate, cost per graduate, non-completion rates, entry qualifications, and initial labour market success are all factors to be measured. These in turn are consolidated into a 'rate of return' which, it is argued, can be used in the future to measure not only differences across the binary line, but also between institutions, and between subjects and departments within institutions. On the research side DES sees measuring outputs as rather harder, for example, in measuring the 'quality' of a book or an article, so it is working on ways of measuring inputs to research and their relationships to teaching.

Performance indicators may in due course prove to be even more controversial than the policy perspective, and it will be interesting to see how the debate about the future of higher education develops over the next year. □

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.